Serials
QH
301
.B615
v.48, no3/4

Vol. 48, Parts 3 and 4 December 1961

PR
THE UNIVERSITY OF MICHIGAN
DEC 28 1961
SCIENCE LIBRARY

BIOMETRIKA

FOUNDED BY
W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

MANAGING EDITOR
E. S. PEARSON

ASSOCIATE EDITOR
J. DURBIN

ISSUED BY
THE BIOMETRIKA OFFICE, UNIVERSITY COLLEGE LONDON

[*Issued December* 1961]

This volume of *Biometrika* is published with the co-operation of

F. N. DAVID — N. L. JOHNSON
M. J. R. HEALY — D. V. LINDLEY
R. L. PLACKETT

A volume containing about 500 pages will be published annually in two half-yearly issues, appearing in June and December.

Papers for publication should be sent to

PROFESSOR E. S. PEARSON
University College London, Gower Street, London, W.C. 1

It is a condition of publication in *Biometrika* that the paper shall not already have been issued elsewhere, and will not be reprinted without leave of the Editors.

Contributors receive 50 copies of their papers free. Joint authors 25 copies each. Order forms for separates are sent to authors with proofs of their papers.

SUBSCRIPTIONS for 1962 (Volume 49)

The subscription price, payable in advance is:
£2. 14*s.* (or **$8.00**) net per volume *including packing and postage.*

Subscribers who buy the journal *Biometrika* for their own and not for general use may apply to the Secretary to be registered as **Personal Subscribers,** for whom the subscription price is **£2. 5*s.*** (payable only in sterling).

Cheques should be made payable to *Biometrika*, crossed 'a/c BIOMETRIKA TRUST' and sent to THE SECRETARY, BIOMETRIKA OFFICE, UNIVERSITY COLLEGE LONDON, GOWER STREET, LONDON, W.C. 1, to whom all orders for Offprints, Index and Table Separates (see pages (iv) to (vi)) should also be sent. For particulars as to the availability of **Back Issues** please see inside back cover.

[Members of the Institute of Mathematical Statistics are reminded that they may subscribe to *Biometrika* through the Institute who will accept subscriptions to *Biometrika* at the special rate of $6.50 if paid to the Treasurer of the Institute before 1 March.]

Biometrika (1961), **48**, 3 *and* 4, *p.* 241
Printed in Great Britain

The two-sample capture-recapture census when tagging and sampling are stratified

BY J. N. DARROCH
Statistical Laboratory, University of Manchester

1. INTRODUCTION AND SUMMARY

1·1. We start by recalling the capture-recapture argument used for the simplest type of experiment with only two samples and negligible death and emigration rates.

Let a animals be taken from a population, marked and put back into it. After allowing time for these a individuals to 'mix' with the others, let a second sample be taken and suppose that it comprises b unmarked individuals and c marked ones. Then, if it is assumed that *every individual has the same probability p of being a member of the second sample*, p is estimated by $\tilde{p} = c/a$ and, if n is the number of unmarked individuals in the population at the time of the second sample, n is estimated by $b/\tilde{p} = ab/c$. We shall denote this estimate by $\tilde{n}_P$ and refer to it as the Peterson estimate, although this name is usually given to

$$\tilde{n}_P + a = a(b+c)/c,$$

the estimate of total population size.

The italicized assumption is both the essence and the weakness of this and all other capture-recapture arguments and is implicit in most work on the subject. (See Darroch (1958, 1959) for a short review of the literature and a discussion of the multiple-recapture census.) In practice it can be violated in many ways which may be summarized as follows:

(i) Animals can differ in their inherent catchability.

(ii) The catchability of an animal may change after being captured and marked.

(iii) The probability p can vary geographically over the region occupied by the population, partly because the animals are more catchable in one locality than another and also because the effort expended in catching them is not uniform over the region.

1·2. In order to cope with (i) and (ii) one would need more than two samples but, by contrast, it is possible to adapt the two-sample experiment to allow for (iii) and the adaptations are the subject of this paper. We shall suppose that, at the time of the second sample, the region occupied by the population is divided into t subregions or strata in each of which p can be assumed uniform. Also that, at the time of the first sample, the region can be divided into s strata where each has the property that when the marked animals are randomly released within it they have the same probability distribution of moving to the t strata. The experimenter is required to use a distinctive mark in each of the s strata in order that he can record the stratum of origin of the recaptured animals.

This subject has been treated in three previous papers known to the author. Most of the problems raised here were first raised by Chapman & Junge (1956) and many of the answers obtained are the same as or similar to theirs, but we shall not attempt a detailed comparison as it would lengthen the present work too greatly. The main difference between their treatment and ours lies in the models used. They estimated the unknown parameters from sets of equations relating them to the expected values of the observed frequencies, and only when

it came to finding error variances did they assume equal probabilities and independence of behaviour for the different animals. We shall start by making the latter assumptions, estimate by the method of maximum likelihood, and then later relax the independence assumptions in an exact parametric fashion.

The earliest of the three papers is by Schaeffer (1951), arising from a tagging experiment which was stratified temporally instead of spatially. Chapman & Junge pointed out that Schaeffer's estimate was not a consistent one but they did not analyse his data. We do so, in § 6.

Beverton & Holt (1957) considered a similar problem but treated it deterministically and assumed that the marked animals are released in only one stratum.

1·3. *Summary.* The underlying assumptions are that all marked animals released in a given first-sample stratum have the same probability distribution of movement to the second-sample strata, and that all animals, marked and unmarked, in a given second-sample stratum have the same probability of capture. These assumptions are sufficient when $s \geqslant t$ but when $s < t$ it is necessary to assume that the unmarked animals move with the same probabilities as the marked animals. As one of the main aims of this paper is to keep the assumptions to a minimum, most of the attention is given to the case $s \geqslant t$ and we look at the case $s < t$ only in § 2·5.

The unknown parameters are the movement probabilities, the catching probabilities, the numbers of unmarked animals present in the different second-sample strata and, in particular, their total. It is usually assumed in two-sample censuses that every marked animal has probability zero of dying or emigrating in the interval between the two samples or, in other words, that the 'survival probability' is one. As this assumption is often not justifiable we have framed the theory in order that it may be avoided if desired. The price paid for dropping this assumption is that, instead of estimating the true (movement and catching) probabilities and true sizes, we have to be content with estimating them after they have been scaled either up or down by the survival probability. However, these scaled parameters do, for the most part, have a useful physical meaning. See §§ 2·1, 2.

Nearly all estimates are found by the method of maximum likelihood and are therefore asymptotically efficient. For $s = t$ the estimates of sizes are given by (8) and (9) and their variances and covariances by (18) and (19). For $s > t$, it is possible to allow the survival probabilities to depend to a certain extent on the stratum of origin. See § 2·4, (8′), (9′), (18′) and (19′).

In § 4 it is shown that if the (hitherto implicit) assumption of *independent* movement and catching of the marked animals is dropped, the estimates are still consistent and the formulae for their variances undergo only minor changes.

In § 5 we examine the validity of using the simple Peterson estimate since it would be unrealistic to assume that the experimenter always knows about the underlying stratification and is able to conduct his experiment to fit in with it. Our conclusions there support most of the previous literature on this question which has been mainly of a heuristic nature.

Finally, in § 6, most of the theory is illustrated on the data of Schaeffer's paper.

2. PROBABILITY MODEL AND ESTIMATION

2·1. Subscript i will refer to the s first-sample strata and subscript j to the t second-sample strata.

Of the a animals tagged, let a_i be released in the ith stratum and, of these, let c_{ij} be caught

in the jth stratum. Let $c_{i.} = \sum_j c_{ij}$, $c_{.j} = \sum_i c_{ij}$. Then $\sum_i c_{i.} = \sum_j c_{.j} = c$. Let θ^*_{ij} denote the probability of moving to the jth stratum for each of the a_i individuals and write $\sum_j \theta^*_{ij} = \phi^*_i$, the survival probability. (The reason for using asterisks will appear in § 2·2.) For the most part, the ϕ^*_i will be assumed equal, with common value ϕ^* say, since this is usually a reasonable biological assumption but, in § 2·4 and § 6·3, they will be estimated on the supposition that they differ in a certain restricted fashion.

Turning to the untagged individuals, we shall assume nothing about their movement (except in § 2·5) and therefore they need not be brought into the picture until the time of the second sample. Accordingly let n^*_j denote the number in the jth stratum at that time and let $n^* = \sum_j n^*_j$. Of the n^*_j let b_j be caught in the second sample.

The fundamental assumption will be that all live individuals in the jth stratum, marked or unmarked, have the same probability, p^*_j say, of being a member of the second sample.

2·2. Assuming independent behaviour as regards movement and catching, we have the following two probability densities.

$$p[\{c_{ij}\}|\{a_i\}] = \frac{\prod_i a_i!}{\prod_i (a_i - c_{i.})! \prod_{ij} c_{ij}!} \prod_i (1 - \sum_j \theta^*_{ij} p^*_j)^{a_i - c_{i.}} \cdot \prod_{ij} (\theta^*_{ij} p^*_j)^{c_{ij}}, \tag{1*}$$

$$p[\{b_j\}] = \prod_j \binom{n^*_j}{b_j} (p^*_j)^{b_j} (1 - p^*_j)^{n^*_j - b_j}. \tag{2*}$$

Thus the likelihood of $\{\theta^*_{ij}\}$, $\{p^*_j\}$, $\{n^*_j\}$ is obtained by multiplying (1*) and (2*), giving $e^{L^*} = e^{L^*_1} \cdot e^{L^*_2}$ say. Differentiating L^* with respect to p^*_j gives

$$\frac{\partial L^*}{\partial p^*_j} = \frac{\partial L^*_1}{\partial p^*_j} + \frac{b_j}{p^*_j} - \frac{n^*_j - b_j}{1 - p^*_j},$$

and maximizing with respect to n^*_j by equating $\Delta_{n^*_j} L^*$ to zero gives

$$\frac{b_j}{p^*_j} = \frac{n^*_j - b_j}{1 - p^*_j}. \tag{3*}$$

Therefore, at the maximum-likelihood value of p^*_j, $\partial L^*/\partial p^*_j = \partial L^*_1/\partial p^*_j$. Also, since θ^*_{ij} is present only in L^*_1, $\partial L^*/\partial \theta^*_{ij} = \partial L^*_1/\partial \theta^*_{ij}$. In other words, maximizing L^* with respect to the unknown parameters is equivalent to maximizing L^*_1 and then (by (3*)) estimating n^*_j by the estimate of b_j/p^*_j. This allows us to make the simplification of speaking of (1*) as *the* likelihood and excluding (2*) from the explicit maximization process.

The next point to be made is that $\{\theta^*_{ij}\}$, $\{p^*_j\}$ must always be non-identifiable to the extent of a multiplicative constant, for their likelihood is the same as that of $\{k\theta^*_{ij}\}$, $\{k^{-1}p^*_i\}$. In particular it is not possible to estimate the survival probability ϕ^*. To tie down this non-identifiability, we shall work with the parameters $\theta_{ij} = \theta^*_{ij}/\phi^*$ and $p_j = \phi^* p^*_j$. These parameters *are* identifiable (provided $s \geqslant t$) and have fairly obvious probability interpretations. Thus, θ_{ij} is the probability of any one of the a_i individuals being in stratum j at the time of the second sample given that it is alive at that time; while p_j is the probability of any animal surviving and, if it is in the jth stratum, of being caught there. Correspondingly, it is not possible to estimate the n^*_j (unless it is known *a priori* that $\phi^* = 1$) and, instead, we estimate $n_j = n^*_j/\phi^*$, $n = \sum_j n_j$. n_j is the number of unmarked individuals in the jth stratum divided

by the probability of survival for the marked individuals. Therefore, assuming that there is no immigration during the experiment and that marked and unmarked individuals have the same probability of survival, n_j is roughly the number of unmarked individuals at the time of the first sample who finish up in the jth stratum and n is roughly the total unmarked population size at the time of the first sample.

To summarize the foregoing, our plan will be to maximize the likelihood of $\{\theta_{ij}\}$, $\{p_j\}$, namely

$$e^L \propto \prod_i (1-\sum_j \theta_{ij}p_j)^{a_i-c_i} \cdot \prod_{ij} (\theta_{ij}p_j)^{c_{ij}}, \tag{1}$$

obtaining estimates $\{\tilde{\theta}_{ij}\}$, $\{\tilde{p}_j\}$ and then to estimate n_j by $\tilde{n}_j = b_j/\tilde{p}_j$ and n by $\tilde{n} = \sum_j \tilde{n}_j$.

The number of parameters in (1) is $st-s$ (for the θ_{ij} subject to the constraints $\sum_j \theta_{ij} = 1$, $i = 1, 2, \ldots, s$) plus t (for the p_j), making a total of

$$N(\theta, p) = st+t-s.$$

Compare this with number of parameters involved when (1) is expressed in terms of $\psi_{ij} = \theta_{ij}p_j$, namely

$$N(\psi) = st.$$

(The point of introducing the ψ system is that it reveals in a simple manner certain important features of the θ, p system.) We see that

$$N(\theta, p) \gtrless N(\psi)$$

according as $s \lesseqgtr t$, and the three cases: $s = t$, $s > t$, $s < t$ will be treated separately.

2·3. When $s = t$, $N(\theta, p) = N(\psi)$ and the two parametric systems are interchangeable provided only that the transformation $\psi_{ij} = \theta_{ij}p_j$ is one-one. In matrix notation it is

$$\boldsymbol{\Psi} = \boldsymbol{\Theta}\mathbf{D}_{\mathbf{p}}, \tag{2}$$

where $\boldsymbol{\Psi} = (\psi_{ij})$, $\boldsymbol{\Theta} = (\theta_{ij})$ and $\mathbf{D}_{\mathbf{p}} = (\delta_{ij}p_j)$, the diagonal matrix whose elements are those of the vector $\mathbf{p} = (p_j)$. Now $\boldsymbol{\Theta}\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the vector of s 1's, and therefore $\boldsymbol{\Psi}\mathbf{D}_{\mathbf{p}}^{-1}\mathbf{1} = \mathbf{1}$. Hence, provided $\boldsymbol{\Psi}$ is non-singular, that is provided $\boldsymbol{\Theta}$ is non-singular,

$$\mathbf{D}_{\mathbf{p}}^{-1}\mathbf{1} = \boldsymbol{\Psi}^{-1}\mathbf{1},$$

or

$$\boldsymbol{\rho} = \boldsymbol{\Psi}^{-1}\mathbf{1}, \tag{3}$$

where $\boldsymbol{\rho} = (\rho_j) = (p_j^{-1})$. (2) and (3) show that the transformation is one-one if and only if $\boldsymbol{\Theta}$ is non-singular.

Maximizing the likelihood with respect to $\{\psi_{ij}\}$ the estimates are

$$\tilde{\psi}_{ij} = c_{ij}/a_i. \tag{4}$$

$\tilde{\psi}_{ij}$ is obviously a consistent estimate of ψ_{ij} as $a_i \to \infty$. Translating into the θ, p system, we deduce that

$$\tilde{\theta}_{ij}\tilde{p}_j = c_{ij}/a_i$$

are the maximum-likelihood equations and give consistent estimates of $\{\theta_{ij}\}$, $\{p_j\}$ provided that $\tilde{\boldsymbol{\Theta}}$ and $\boldsymbol{\Theta}$ are non-singular. From (4) it follows that $\tilde{\boldsymbol{\Psi}}$ (and therefore also $\tilde{\boldsymbol{\Theta}}$) is non-singular if and only if $\mathbf{C} = (c_{ij})$ is. If $\mathbf{C}$ is non-singular then, from (3) and (4),

$$\tilde{\boldsymbol{\rho}} = \mathbf{C}^{-1}\mathbf{D}_{\mathbf{a}}\mathbf{1} = \mathbf{C}^{-1}\mathbf{a}, \tag{5}$$

and since $\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C}$ converges in probability to $\boldsymbol{\Theta}\mathbf{D}_{\mathbf{p}}$ as the $a_i \to \infty$, $\tilde{\boldsymbol{\rho}}$ converges to

$$\mathbf{D}_{\mathbf{p}}^{-1}\boldsymbol{\Theta}^{-1}\mathbf{1} = \mathbf{D}_{\mathbf{p}}^{-1}\mathbf{1} = \boldsymbol{\rho}$$

provided $\boldsymbol{\Theta}$ is non-singular (as already stated). On the other hand, if $\boldsymbol{\Theta}$ is singular, $\tilde{\boldsymbol{\rho}}$ diverges. It is easily seen that the elements of $\tilde{\boldsymbol{\rho}} = \mathbf{C}^{-1}\mathbf{a}$ are not necessarily positive and greater than one. In other words, it is quite possible to obtain a $\tilde{p}_j$ which does not lie in the interval $[0, 1]$. The first diagnosis of this anomaly that springs to mind is that the estimation procedure should have somehow incorporated the restriction that all p_j (and all θ_{ij}) lie in $[0, 1]$. Usually when estimating probabilities there is no need to do this explicitly as their estimates turn out to be proportions and, indeed, as far as the ψ_{ij} are concerned, this is what happens in the present case. It is only in transferring from $\tilde{\psi}_{ij} = c_{ij}/a_i$ to $\tilde{\theta}_{ij}$ and $\tilde{p}_j$ that the anomaly occurs. As this transfer is made by imposing the constraints $\sum_j \tilde{\theta}_{ij} = 1$ $(i = 1, 2, \ldots, s)$ it is these that are at the root of the trouble and there are, broadly speaking, two possible diagnoses. Namely, (i) the true ϕ_i^* are not equal or (ii) the true ϕ_i^* are equal but sampling errors are overwhelmingly large. (i) can be allowed for by reducing the number of second-sample strata and applying the theory of § 2·4. This is further discussed in the case of the numerical example in §§ 6·2, 3. (ii) means that the $\tilde{p}_j$ and the $\tilde{n}_j = b_j/\tilde{p}_j$ are virtually useless as estimates of p_j and n_j and this would no doubt be confirmed by their having large variances. However, this does not necessarily mean that the variance of $\tilde{n} = \sum_j \tilde{n}_j$ is also large as the covariances of the $\tilde{n}_j$ may well be large and negative. Again, see the discussion of the numerical example in § 6·3.

When $\mathbf{C}$ is singular the likelihood equations in the ψ system are invalid and we must look at those in the θ, p system. $\partial L/\partial p_j = 0$ gives

$$\sum_i \frac{(a_i - c_{i\cdot})\,\theta_{ij} p_j}{1 - \sum_j \theta_{ij} p_j} = c_{\cdot j}. \tag{6}$$

Introducing s Lagrangian multipliers $\{\lambda_i\}$ corresponding to the constraints $\sum_j \theta_{ij} = 1$, $\partial L/\partial \theta_{ij} = 0$ gives

$$\frac{(a_i - c_{i\cdot})\,\theta_{ij} p_j}{1 - \sum_j \theta_{ij} p_j} = c_{ij} + \lambda_i \theta_{ij}. \tag{7}$$

These equations have no easy solution except when $s = t = 2$. (They do, of course, have an easy solution when $\mathbf{C}$ is non-singular. For (6) and (7) imply that $\sum_i \lambda_i \tilde{\theta}_{ij} = 0$ all j, which, if $\tilde{\boldsymbol{\Theta}}$ is non-singular, implies that $\lambda_i = 0$ all i and hence that $\tilde{\theta}_{ij}\tilde{p}_j = c_{ij}/a_{i\cdot}$.)

It is obvious from the foregoing that the experimenter must try and avoid stratifications which make $\boldsymbol{\Theta}$ singular; and the further $\boldsymbol{\Theta}$ is from singularity the more reliable are the estimates $\{\tilde{\theta}_{ij}\}, \{\tilde{p}_j\}$, a fact which will be confirmed by the variance formulae in § 3. There are a few trivial instances of singularity which may be foreseeable. For example, if $\theta_{ij} = 0$ all i, the jth stratum is effectively non-existent. If $\theta_{ij} = \theta_{lj}$ for all j, the ith and lth strata can be combined. If $\theta_{ij}/\theta_{ik} = \omega$, independent of i, the jth and kth strata can be combined into a Jth, say, with $\theta_{iJ} = \theta_{ij} + \theta_{ik}$, $p_J = (\omega p_j + p_k)/(\omega + 1)$. But it will sometimes be less easy to avoid the more general instances of singularity.

If $\boldsymbol{\Theta}$ is non-singular and the sample sizes are large, a singular $\mathbf{C}$ will be extremely improbable. But, if it does occur, then, because of the difficulty of solving the likelihood equations and in view of the fact that the variance formulae of § 3 are derived conditional on $\mathbf{C}$ being non-singular, the best procedure will be to group two of the first-sample strata and two of the second-sample strata in such a way as to eliminate the singularity. (The

first-sample grouping may be omitted, making $s > t$ and involving us in the theory of § 2·4.) If both groupings are made when $s = t = 2$ the result is an unstratified analysis and it is interesting that the maximum-likelihood solution with λ_1 and λ_2 non-zero throws out strong hints that this is the correct procedure. For the latter solution is

$$\tilde{\theta}_{11} = \tilde{\theta}_{21} = [(a_1c_{21} - a_2c_{11}) + (c_{11}c_{22} - c_{12}c_{21})]/(a_1c_{2\cdot} - a_2c_{1\cdot}),$$

$$\tilde{\theta}_{12} = \tilde{\theta}_{22} = 1 - \tilde{\theta}_{11}, \quad \tilde{p}_1 = c_{\cdot 1}/a\tilde{\theta}_{11}, \quad \tilde{p}_2 = c_{\cdot 2}/a\tilde{\theta}_{22}.$$

(This solution, which always exists, is obviously not consistent.) When $c_{11}c_{22} - c_{12}c_{21} = 0$, besides $\tilde{\theta}_{11} = \tilde{\theta}_{21}$ and $\tilde{\theta}_{12} = \tilde{\theta}_{22}$ which indicates a combination of the first-sample strata, it turns out that $\tilde{p}_1 = \tilde{p}_2 = c/a$ which indicates a combination of the second-sample strata.

Finally, let us use (5) to estimate $\mathbf{n} = (n_j)$. Since $\tilde{\mathbf{n}} = (b_j/\tilde{p}_j) = \mathbf{D_b}\tilde{\boldsymbol{\rho}}$,

$$\tilde{\mathbf{n}} = \mathbf{D_b}\mathbf{C}^{-1}\mathbf{a}, \tag{8}$$

first found by Chapman & Junge under similar assumptions. Also, since $\tilde{n} = \mathbf{1}'\tilde{\mathbf{n}}$,

$$\tilde{n} = \mathbf{b}'\mathbf{C}^{-1}\mathbf{a}. \tag{9}$$

2·4. When $s > t$, $N(\theta, p) < N(\psi)$ and the ψ_{ij} are mathematically dependent functions of $\{\theta_{ij}\}, \{p_j\}$. This means that the simple estimation equations $\tilde{\psi}_{ij} = c_{ij}/a_i$ are invalid and we have to solve (6) and (7). These equations have no explicit solution (except in the trivial case $t = 1$) and, for any given data, they would therefore have to be solved iteratively and the information matrix inverted to find error variances and covariances. If s and t are not small this is not an attractive proposition but, fortunately, it is easily avoided as follows.

We continue to make $N(\theta, p) = N(\psi)$ by allowing the ϕ_i^* to differ. Let $\overline{\phi}^* = (\sum_i \phi_i^*)/s$ and let us work with parameters

$$\theta_{ij} = \theta_{ij}^*/\overline{\phi}^*, \quad \phi_i = \sum_j \theta_{ij}, \quad p_j = \overline{\phi}^* p_j^*.$$

These parameters correspond closely to the θ_{ij} and p_j of § 2·2 but do not, unfortunately, have simple probability interpretations. This gives a total of $st - 1 + t$ parameters, the subtraction of one being due to $\sum_{ij} \theta_{ij} = s$. Instead of imposing $s - 1$ constraints: $\phi_1 = \phi_2 = \ldots = \phi_s$, as we effectively did in § 2·3, let us instead only impose $t - 1$. Then $N(\theta, p) = st$ and (with certain provisos) the estimation equations are $\tilde{\theta}_{ij}\tilde{p}_j = c_{ij}/a_i$. The $t - 1$ (independent) constraints will usually take the form $\phi_i - \phi_k = 0$ or possibly $\phi_i + \phi_k - 2\phi_1 = 0$ but we can write them generally as

$$\sum_i u_{ki}\phi_i = 0 \quad (k = 1, 2, \ldots, t-1).$$

Also

$$\sum_i u_{ti}\phi_i = 1$$

if we define $u_{ti} = 1/s$. The matrix $\mathbf{U} = (u_{ki})\,(k = 1, 2, \ldots, t)$, $(i = 1, 2, \ldots, s)$ is of full rank t because the first $t - 1$ rows are independent among themselves and, provided they make sense as constraints, the last row is independent of them. For, if not, the first $t - 1$ rows could be combined to give $\sum_i \phi_i = 0$, which is nonsensical.

Let $\mathbf{v}'$ denote the $1 \times t$ vector $(0, 0, \ldots, 0, 1)$. Then

$$\mathbf{U}\boldsymbol{\Theta}\mathbf{1} = \mathbf{v}$$

replacing $\boldsymbol{\Theta}\mathbf{1} = \mathbf{1}$ of § 2·3. Adapting the argument of § 2·3 we find that, provided $\mathbf{U}\boldsymbol{\Theta}$ and $\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C}$ are non-singular or, in other words, provided $\boldsymbol{\Theta}$ and $\mathbf{C}$ are of full rank t, consistent estimates of $\boldsymbol{\rho} = (p_j^{-1})$ and $\boldsymbol{\Theta}$ are given by

$$\tilde{\boldsymbol{\rho}} = (\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C})^{-1}\mathbf{v}, \tag{5'}$$

$$\tilde{\boldsymbol{\Theta}} = \mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C}\mathbf{D}_{\boldsymbol{\rho}}.$$

In the trivial case $t = 1$, the $\phi_i = \theta_{i1}$ are estimated without constraints giving

$$\tilde{p}_1 = (\sum_i c_{i1}/a_i)/s, \quad \tilde{\phi}_i = c_{i1}/a_i\tilde{p}_1.$$

Finally, $\mathbf{n} = (n_j) = (n_j^*/\overline{\phi}^*)$ is estimated by

$$\tilde{\mathbf{n}} = \mathbf{D}_{\mathbf{b}}(\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C})^{-1}\mathbf{v} \tag{8'}$$

and $n = n^*/\overline{\phi}^*$ by

$$\tilde{n} = \mathbf{b}'(\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C})^{-1}\mathbf{v}. \tag{9'}$$

If $\overline{\phi}^*$ is also the 'average' probability of survival of the unmarked individuals, then n_j and n have the same interpretations as in § 2·3. It is of course not necessary to work with $\overline{\phi}^* = (\sum_i \overline{\phi}_i^*)/s$ and any linear combination $\sum_i \nu_i \phi_i^*$, where $\sum_i \nu_i = 1$, would serve instead, and would only involve changing u_{ii} from $1/s$ to ν_i.

2·5. When $s < t$, $N(\theta, p) > N(\psi)$ and $\{\theta_{ij}\}, \{p_j\}$ are not identifiable; only $N(\psi)$ functions of them are. If we try to estimate $\{p_j\}$ from the equations $\tilde{\psi}_{ij} = c_{ij}/a_i$ (or, equivalently, from (6) and (7)) we only get s equations: $\sum_j c_{ij}\tilde{p}_j^{-1} = a_i$, for t unknowns. Since $\tilde{p}_j$ cannot be determined, neither can $\tilde{n}_j = b_j/\tilde{p}_j$.

At this point it is appropriate to consider the effect of assuming that marked and unmarked animals have the same movement pattern since, with this assumption, n can be estimated even though the n_j cannot. The unmarked individuals are now brought into the picture at the time of the first sample by letting m_i denote the number in the ith stratum. The density $p[\{b_j\}]$ becomes a convolution of s multinomial densities, the ith having parameters $m_i, \psi_{i1}, \psi_{i2}, \ldots, \psi_{it}$, and the likelihood equations are intractable. The best we can do, therefore, is to consider the moment equations

$$\sum_i m_i\psi_{ij} = b_j,$$

$$a_i\psi_{ij} = c_{ij}.$$

From these we get t equations

$$\sum_i m_i c_{ij}/a_i = b_j \tag{10}$$

for s unknown parameters $\{m_i\}$. When $s = t$ the solution of (10) is

$$\tilde{\mathbf{m}}' = \mathbf{b}'\mathbf{C}^{-1}\mathbf{D}_{\mathbf{a}},$$

giving

$$\sum_i \tilde{m}_i = \mathbf{b}'\mathbf{C}^{-1}\mathbf{a} \tag{11}$$

which is equivalent to (9). When $s < t$ there are more than enough equations and two courses are open: either equations (10) may be replaced by s linear combinations of them, or the number of parameters can be increased by relaxing the equality of some of the movement probabilities or by introducing some immigration parameters. We shall not go into details.

Finally we note that in deriving (10), we used ψ_{ij} merely as the probability of being caught in the jth stratum for an individual starting in the ith stratum and did not have to

express it as the product of a movement probability and a catching probability. Consequently, if it is expressed in this form, it can be taken as $\theta_{ij}p_{ij}$ with no restriction on the relative values of the $\phi_i(= \sum_j \theta_{ij})$. Thus (as pointed out by Chapman & Junge), in return for imposing the assumption that marked and unmarked animals have the same movement pattern, we can allow the survival probabilities and the probabilities of capture to depend on the stratum of origin. However, we consider that this involves losing more on the roundabouts than is gained on the swings and we shall therefore keep to our original assumptions for the remainder of the paper, and merely note that much of the ensuing theory for $s \geqslant t$ could be adapted under the different assumptions for $s \leqslant t$.

3. First and second moments of the estimates

3·1. We first find the approximate bias of $\tilde{\boldsymbol{\rho}}$ for the limit process: $\{\theta_{ij}\}$, $\{p_j\}$ constant and all $a_i \to \infty$ in such a way that a_i/a is fixed. Both $\boldsymbol{\Theta}$ and $\mathbf{C}$ will be assumed to be of rank t, and we shall neglect the probability that $\mathbf{C}$ is of rank less than t.

Let
$$\boldsymbol{\Gamma} = E[\mathbf{C}] = \mathbf{D}_a \boldsymbol{\Theta} \mathbf{D}_p$$
and
$$\mathbf{C} - \boldsymbol{\Gamma} = \mathbf{Z} = (z_{ij}).$$

When $s = t$, $\tilde{\boldsymbol{\rho}} = \mathbf{C}^{-1}\mathbf{a}$ and

$$\mathbf{C}^{-1} = [\mathbf{I} - \boldsymbol{\Gamma}^{-1}\mathbf{Z} + (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^2 - \ldots + (-1)^{m-1}(\boldsymbol{\Gamma}^{-1}\mathbf{Z})^{m-1}]\boldsymbol{\Gamma}^{-1} + (-1)^m(\boldsymbol{\Gamma}^{-1}\mathbf{Z})^m\mathbf{C}^{-1}.$$

Therefore, since $\boldsymbol{\Gamma}^{-1}\mathbf{a} = \boldsymbol{\rho}$, we have

$$E[\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho}] = E[-(\boldsymbol{\Gamma}^{-1}\mathbf{Z}) + (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^2 - \ldots + (-1)^{m-1}(\boldsymbol{\Gamma}^{-1}\mathbf{Z})^{m-1}]\boldsymbol{\rho} + \mathbf{r},$$

where $\mathbf{r} = (r_j) = (-1)^m E[(\boldsymbol{\Gamma}^{-1}\mathbf{Z})^m \mathbf{C}^{-1}\mathbf{a}]$, the remainder vector. Let $\mathbf{F} = (f_{ij}) = (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^m$ and let C_{ij} denote the co-factor of c_{ij} in $\mathbf{C}$. Then, since $|C|$, the determinant of $\mathbf{C}$, is assumed numerically greater than or equal to one,

$$|r_j| = |E[\sum_{k,l} f_{jk}C_{lk}a_l/|C|]| \leqslant \sum_{k,l} |E[f_{jk}C_{lk}]|\, a_l \leqslant \sum_{k,l} (E[f_{jk}^2]\, E[C_{lk}^2])^{\frac{1}{2}} a_l.$$

Now $E[f_{jk}^2]$ is the sum of products of terms of $\boldsymbol{\Gamma}^{-1}$ of total power $2m$ and moments of $\mathbf{Z}$ of order $2m$. The former are $O(a^{-2m})$ and the latter, since they are multinomial moments, are $O(a^m)$. Also, C_{lk} is numerically less than $(\prod_i c_{i.})/c_{l.}$ which is $O(a^{s-1})$. Therefore

$$|r_j| \leqslant O(a^{-m+\frac{1}{2}m+s-1+1}) = O(a^{-\frac{1}{2}m+s}) \to 0$$

for $m > 2s$. It is therefore permissible to write $E[\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho}]$ as the infinite series

$$E[-(\boldsymbol{\Gamma}^{-1}\mathbf{Z}) + (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^2 - (\boldsymbol{\Gamma}^{-1}\mathbf{Z})^3 + \ldots]\boldsymbol{\rho}.$$

The first term is zero since $E[\mathbf{Z}] = \mathbf{O}$, the $(2m)$th is $O(a^{-m})$ and the $(2m+1)$th is $O(a^{-m-1})$, $m \geqslant 1$.

Therefore the bias of $\tilde{p}_j = 1/\tilde{p}_j$ is $O(a^{-1})$ and this is asymptotically negligible compared with its root mean square which we shall show in § 3·2 to be $O(a^{-\frac{1}{2}})$. The same holds true, when $s > t$, of $\tilde{\boldsymbol{\rho}} = (\mathbf{U}\mathbf{D}_a^{-1}\mathbf{C})^{-1}\mathbf{v}$.

Next, let us augment the above limit process by supposing that all $n_j \to \infty$ in such a way that n_j/n and a/n remain fixed. Let $\boldsymbol{\beta} = E[\mathbf{b}]$. As $\mathbf{b}$ is independent of $\mathbf{C}$,

$$E[\hat{\mathbf{n}} - \mathbf{n}] = \mathbf{D}_{\boldsymbol{\beta}} E[\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}], \quad E[\tilde{n} - n] = \boldsymbol{\beta}' E[\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho}]$$

which are both $O(1)$ and negligible compared with the root mean squares which are $O(n^{\frac{1}{2}})$.

3·2. Because the bias of $\tilde{\boldsymbol{\rho}}$ is negligible, its variance-covariance matrix is

$$E[(\tilde{\boldsymbol{\rho}}-\boldsymbol{\rho})(\tilde{\boldsymbol{\rho}}-\boldsymbol{\rho})'] = \boldsymbol{\Sigma}$$

say, and, arguing as in § 3·1, this matrix can be expanded as a doubly infinite series whose terms are $O(a^{-m})$, $m \geqslant 1$. Suppose $s = t$. Then the leading term, of size $O(a^{-1})$, is

$$E[\boldsymbol{\Gamma}^{-1}\mathbf{Z}\boldsymbol{\rho}\boldsymbol{\rho}'\mathbf{Z}'\boldsymbol{\Gamma}'^{-1}]. \tag{12}$$

The (i, l) element of $E[\mathbf{Z}\boldsymbol{\rho}\boldsymbol{\rho}'\mathbf{Z}']$ is

$$E[\sum_{jk} z_{ij}\rho_j\rho_k z_{lk}]. \tag{13}$$

Now
$$E[z_{ij}z_{lk}] = \delta_{il}a_i[\delta_{jk}\theta_{ij}p_j - \theta_{ij}p_j\theta_{ik}p_k] \tag{14}$$

since c_{ij} and c_{lk} are independent when $i \neq l$ and, when $i = l$, have a multinomial distribution. Substituting (14) in (13) we get

$$\delta_{il}a_i[\sum_j \theta_{ij}/p_j - (\sum_j \theta_{ij})(\sum_k \theta_{ik})] = \delta_{il}a_i[\sum_j \theta_{ij}/p_j - 1]$$

and this is the (i, l) element of $\mathbf{D}_{\mathbf{a}}\mathbf{D}_{\boldsymbol{\mu}}$, where $\boldsymbol{\mu} = (\mu_i)$ and

$$\mu_i = \sum_j \theta_{ij}/p_j - 1. \tag{15}$$

Substituting in (13), we can say that

$$\boldsymbol{\Sigma} \sim \boldsymbol{\Gamma}^{-1}\mathbf{D}_{\mathbf{a}}\mathbf{D}_{\boldsymbol{\mu}}\boldsymbol{\Gamma}'^{-1} = \mathbf{D}_{\boldsymbol{\rho}}\boldsymbol{\Theta}^{-1}\mathbf{D}_{\boldsymbol{\mu}}\mathbf{D}_{\mathbf{a}}^{-1}\boldsymbol{\Theta}'^{-1}\mathbf{D}_{\boldsymbol{\rho}}. \tag{16}$$

Notice that, when $\boldsymbol{\Theta}$ is near to singularity and the elements of $\boldsymbol{\Theta}^{-1}$ are large, $\tilde{\boldsymbol{\rho}}$ will be an inaccurate estimate of $\boldsymbol{\rho}$.

When $s > t$, (16) is changed to

$$\boldsymbol{\Sigma} \sim \mathbf{D}_{\boldsymbol{\rho}}(\mathbf{U}\boldsymbol{\Theta})^{-1}\mathbf{U}\mathbf{D}_{\boldsymbol{\mu}}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{U}'(\mathbf{U}\boldsymbol{\Theta})'^{-1}\mathbf{D}_{\boldsymbol{\rho}}, \tag{16'}$$

where
$$\mu_i = \sum_j \theta_{ij}/p_j - \phi_i^2. \tag{15'}$$

3·3. Let $\mathbf{y}$ denote the deviation of $\mathbf{b}$ from its expected value $\boldsymbol{\beta} = \mathbf{D}_{\boldsymbol{\rho}}\mathbf{n}$. Then

$$\tilde{\mathbf{n}} - \mathbf{n} = \mathbf{D}_{\mathbf{b}}\tilde{\boldsymbol{\rho}} - \mathbf{D}_{\boldsymbol{\beta}}\boldsymbol{\rho} = \mathbf{D}_{\boldsymbol{\beta}}(\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho}) + \mathbf{D}_{\mathbf{y}}\tilde{\boldsymbol{\rho}}.$$

Therefore
$$E[(\tilde{\mathbf{n}} - \mathbf{n})(\tilde{\mathbf{n}} - \mathbf{n})'] = \mathbf{D}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}\mathbf{D}_{\boldsymbol{\beta}} + E[\mathbf{D}_{\mathbf{y}}\tilde{\boldsymbol{\rho}}\tilde{\boldsymbol{\rho}}'\mathbf{D}_{\mathbf{y}}]. \tag{17}$$

The leading term in (17) is $O(n)$ $(= O(a))$ and the others are $O(n^{-m})$, $m \geqslant 0$. Retaining only the former, the second term in (17) may be replaced by $E[\mathbf{D}_{\mathbf{y}}\boldsymbol{\rho}\boldsymbol{\rho}'\mathbf{D}_{\mathbf{y}}] = \mathbf{D}_{\mathbf{n}}[\mathbf{D}_{\boldsymbol{\rho}} - \mathbf{I}\phi^{*-1}]$ since $E[y_j y_k] = \delta_{jk}n_j p_j(1 - p_j\phi^{*-1})$. Substituting (16) when $s = t$

$$E[(\tilde{\mathbf{n}} - \mathbf{n})(\tilde{\mathbf{n}} - \mathbf{n})'] \sim \mathbf{D}_{\mathbf{n}}\boldsymbol{\Theta}^{-1}\mathbf{D}_{\boldsymbol{\mu}}\mathbf{D}_{\mathbf{a}}^{-1}\boldsymbol{\Theta}'^{-1}\mathbf{D}_{\mathbf{n}} + \mathbf{D}_{\mathbf{n}}(\mathbf{D}_{\boldsymbol{\rho}} - \mathbf{I}\phi^{*-1}). \tag{18}$$

Since $(\tilde{n} - n)^2 = \mathbf{1}'(\tilde{\mathbf{n}} - \mathbf{n})(\tilde{\mathbf{n}} - \mathbf{n})'\mathbf{1}$,

$$\begin{aligned} E[(\tilde{n} - n)^2] &\sim \mathbf{n}'\boldsymbol{\Theta}^{-1}\mathbf{D}_{\boldsymbol{\mu}}\mathbf{D}_{\mathbf{a}}^{-1}\boldsymbol{\Theta}'^{-1}\mathbf{n} + \mathbf{n}'(\boldsymbol{\rho} - \mathbf{1}\phi^{*-1}) \\ &= \sum_i \eta_i^2\mu_i/a_i + \sum_j n_j(\rho_j - \phi^{*-1}), \end{aligned} \tag{19}$$

where
$$\boldsymbol{\eta}' = (\eta_i) = \mathbf{n}'\boldsymbol{\Theta}^{-1}. \tag{20}$$

If $\boldsymbol{\Theta}$ is near to singularity the low accuracy of $\tilde{\boldsymbol{\rho}}$ impairs that of $\tilde{\mathbf{n}} = (\tilde{n}_j)$ as is to be expected, but the accuracy of $\tilde{n} = \sum_j \tilde{n}_j$ is almost unimpaired. We have so far specifically avoided any assumption about the movement of the unmarked individuals between the two samples but,

to explain the last remark, suppose that they have the same movement pattern as the unmarked individuals. Suppose, also, that there is no immigration and let m_i denote the number of individuals in the ith stratum. There are two ways of formally equating the movement patterns. One way is to think of n_j^* as a random variable with expectation $\sum_i m_i \theta_{ij}^*$, but this involves some tiresome and unimportant complications. Instead, we shall continue to think of n_j^* as a parameter by the device of *defining* $\{\theta_{ij}^*\}$ by the movement of the unmarked animals. Then $\sum_i m_i \theta_{ij}^* = n_j^*$ or $\sum_i m_i \theta_{ij} = n_j$ or

$$\mathbf{m}'\boldsymbol{\Theta} = \mathbf{n}'. \tag{21}$$

Thus, if the movement patterns are the same, $\boldsymbol{\eta} = \mathbf{m}$. The only point we wish to establish from these considerations is that, even if the movement patterns are different, the η_i are likely to be positive and, since $\sum_i \eta_i = n$, the fact that $\boldsymbol{\Theta}$ is near to singularity has no effect on (19). On the other hand, if $\mathbf{n}$ and $\boldsymbol{\Theta}$ are 'incompatible' in the sense that some of the η_i are negative, either because of very different movement patterns or because of immigrants increasing the numbers of individuals in the t strata disproportionately, this does increase $E[(\tilde{n}-n)^2]$ because the coefficient of η_i^2 in (19) is positive. Of course, these conclusions only apply to the leading terms of size $O(n)$ in $E[(\tilde{n}-n)^2]$. They do not apply to the terms of smaller order and the latter might have to be taken into account if $\boldsymbol{\Theta}$ was very near to singularity.

When $s > t$,

$$E[(\tilde{n}-n)(\tilde{\mathbf{n}}-\mathbf{n})'] \sim \mathbf{D}_{\mathbf{n}}(\mathbf{U}\boldsymbol{\Theta})^{-1}\mathbf{U}\mathbf{D}_{\boldsymbol{\mu}}\mathbf{U}'(\mathbf{U}\boldsymbol{\Theta})'^{-1}\mathbf{D}_{\mathbf{n}} + \mathbf{D}_{\mathbf{n}}(\mathbf{D}_{\boldsymbol{\rho}} - \mathbf{I}\overline{\phi}^{*-1}) \tag{18'}$$

and

$$E[(\tilde{n}-n)^2] \sim \sum_i \eta_i^2 \mu_i / a_i + \sum_j n_j(\rho_j - \overline{\phi}^{*-1}), \tag{19'}$$

where

$$\boldsymbol{\eta}' = \mathbf{n}'(\mathbf{U}\boldsymbol{\Theta})^{-1}\mathbf{U}. \tag{20'}$$

From (20′) we have $\boldsymbol{\eta}'\boldsymbol{\Theta} = \mathbf{n}'$ and, if the movements of the marked and unmarked animals are the same, $\mathbf{m}'\boldsymbol{\Theta} = \mathbf{n}'$. However, besides $\mathbf{m}$, there is now an infinity of vectors $\boldsymbol{\xi}$ such that $\boldsymbol{\xi}'\boldsymbol{\Theta} = \mathbf{n}'$ and we cannot infer that $\boldsymbol{\eta} = \mathbf{m}$. Therefore, the above remarks for $s = t$ cannot be extended for $s > t$.

Finally, $\tilde{n}$ is a consistent estimate of n in the sense that, for the limit process we are using, $\tilde{n}/n$ converges in probability to one.

4. 'CONTAGIOUS' MOVEMENT AND CATCHING

4·1. In writing $p[\{c_{ij}\}|\{a_i\}]$ as a product of multinomial densities (see (1)) it was assumed that the a_i individuals released in the ith stratum move and are caught or not caught (i) independently of those released in any other stratum, and (ii) independently of each other. (i) is a very reasonable assumption, but (ii) is less likely to hold true in practice and in this section we examine the consequences of relaxing it.

Let us refer to the a_i individuals as $I_1, I_2, \ldots, I_{a_i}$. In considering how, for instance, the 'experimental histories' of I_α and I_β may be dependent we must distinguish between the two stages: movement to one of the t strata and, once there, being caught or not caught. To assist this distinction, let a_{ij} denote the number of individuals released in the ith stratum who are alive in the jth at the time of the second sample. Then $a_i - \sum_j a_{ij}$ is the number who die or emigrate between the two samples. We shall first find the distribution of the a_{ij} and then the

conditional distribution of c_{ij} given a_{ij} and, in doing this, it is convenient to revert to the original parameters $\{\theta^*_{ij}\}$, $\{p^*_j\}$ which are now interpreted as common *marginal* probabilities, that is the probabilities for any single individual disregarding the histories of the other individuals.

We note that, whatever form the dependence takes,

$$E[c_{ij}] = E[E[c_{ij}|a_{ij}]] = E[a_{ij}p^*_j] = a_i\theta^*_{ij}p^*_j$$

as in § 2 and therefore the $\tilde{\rho}_j$ are still consistent provided their variances still tend to zero as $a_i \to \infty$. It is these variances in which we shall be mainly interested, and since they principally depend on var (c_{ij}), cov (c_{ij}, c_{ik}) we must find how the latter are changed by dependence from $a_i\theta^*_{ij}p^*_j(1-\theta^*_{ij}p^*_j)$ and $-a_i\theta^*_{ij}p^*_j\theta^*_{ik}p^*_k$, respectively.

4·2. To construct $p[\{a_{ij}\}]$ it is necessary to specify how such probabilities as

$$P[I_\alpha \to j | I_\beta \to k, I_\gamma \to l]$$

(using a self-explanatory notation) differ from θ^*_{ij}. The assumption that there is no difference is really justified only if $I_1, I_2, \ldots, I_{a_i}$ are each released at a random point in the stratum and this would be difficult to achieve in practice. Two possibilities which are more likely are that they are released either close together in a randomly chosen subarea or over a carefully spaced grid of points. In the former case we should expect that, for instance

$$P[I_\alpha \to j | I_\beta \to j, I_\gamma \to j] \geqslant P[I_\alpha \to j | I_\beta \to j] \geqslant \theta^*_{ij},$$

and in the latter case a reversal of these inequalities. In other words, if the animals start close together, they are more likely than otherwise to move to the same stratum whereas, if they start with maximum possible distances between them, they are more likely than otherwise to move to different strata. The point of release may be thought of as determining the initial movement of an animal to the extent that the latter depends not only on such things as topography, wind or current, temperature, the local food distribution, but also, especially if the species is gregarious, on the location and movement of the nearest group of unmarked animals. (It should be noted that we are fortunately not required to try and describe the very complex interdependence of the movements of the unmarked animals nor the dependence of the marked on the unmarked, only the interdependence of the marked.)

Let us call the above two types of dependence 'positive' and 'negative', respectively. Then, as a first approximation, we may describe them in terms of the statistical concept of *contagion* by a simple generalization of Polya's urn model. Since death is a possible contingency it is reasonable to assume that it is contagious also and it can be thought of formally as the $(t+1)$th stratum, letting $a_{i,t+1} = a_i - \sum_j a_{ij}$ and $\theta^*_{i,t+1} = 1-\phi^*_i$.

Consider an urn containing f_i balls of which f_{ih} are marked 'h' and where $f_{ih}/f_i = \theta^*_{ih}$ $(h = 1, 2, \ldots, t+1)$. Thus the proportions of balls in the urn are the marginal probabilities. If there is positive contagion, the conditional probability $P[I_r \to h_r \,|\, I_1 \to h_1, \ldots, I_{r-1} \to h_{r-1}]$ is the proportion marked 'h_r' after adding g_i marked 'h_1', g_i marked 'h_2', ..., g_i marked 'h_{r-1}'. In this way we obtain

$$P[I_r \to h_r, r = 1, 2, \ldots, a_i] = \prod_{h+1}^{t+1} \frac{(f_{ih}/g_i + a_{ih} - 1)!}{(f_{ih}/g_i - 1)!} \Big/ \frac{(f_i/g_i + a_i - 1)!}{(f_i/g_i - 1)!}.$$

(If unequal numbers $g_{ih_1}, g_{ih_2}, \ldots, g_{ih_{r-1}}$ of balls are added, the resulting density is very difficult to handle and, moreover, depends on the ordering of the individuals, a feature we do not desire.) Hence

$$P[\{a_{ih}\}] = \prod_{h=1}^{t+1} \binom{f_{ih}/g_i + a_{ih} - 1}{a_{ih}} \Big/ \binom{f_i/g_i + a_i - 1}{a_i}. \tag{22}$$

It is readily deduced from (22) that

$$\text{var}\,(a_{ih}) = K_i a_i \theta^*_{ih}(1-\theta^*_{ih}), \quad \text{cov}\,(a_{ih}, a_{ik}) = -K_i a_i \theta^*_{ih}\theta^*_{ik}, \tag{23}$$

where $K_i = (f_i/g_i + a_i)/(f_i/g_i + 1)$. Thus $1 \leqslant K \leqslant a_i$, the extreme values corresponding respectively to no contagion and identical behaviour of the a_i individuals.

If there is negative contagion, balls are subtracted instead of added and provided $f_i > g_i a_i$, an obviously necessary condition, (23) still holds with $K_i = (f_i/g_i - a_i)/(f_i/g_i - 1)$. Thus $K_i \leqslant 1$ and, in practice, would be substantially greater than zero.

4·3. If the catching in the jth stratum is not uniform but is concentrated in one or more subareas and if, having a common stratum of origin, the a_{ij} individuals are not uniformly distributed in the jth stratum, there will be a certain amount of positive dependence in their being caught or not caught. Otherwise, it is fair to assume independence. (We shall ignore the possible negative dependence of the catching of individuals from different strata of origin.) A density analogous to (22) can be constructed for $p[c_{ij}|a_{ij}]$ and

$$E[(c_{ij} - a_{ij}p^*_j)^2|a_{ij}] = L_{ij}a_{ij}p^*_j(1-p^*_j), \tag{24}$$

where $1 \leqslant L_{ij} \leqslant a_{ij}$.

4·4. In the following we can generalize from specifically contagious dependence as described by urn models to any form of dependence which alters the variances and covariances in the simple manner given by (23) and (24), where $K_i \gtrless 1$ and $L_{ij} \geqslant 1$ (but probably not much greater than one). Combining (23) and (24),

$$\text{var}\,(c_{ij}) = L_{ij}a_i\theta^*_{ij}p^*_j(1-p^*_j) + K_i a_i \theta^*_{ij}(1-\theta^*_{ij})p^{*2}_j,$$
$$\text{cov}\,(c_{ij}, c_{ik}) = -K_i a_i \theta^*_{ij}p^*_j\theta^*_{ik}p^*_k.$$

Suppose $s = t$. Then using these new variances and covariances in place of (14), we easily find that it is still true that

$$\boldsymbol{\Sigma} \sim \mathbf{D}_\rho \boldsymbol{\Theta}^{-1}\mathbf{D}_\mu \mathbf{D}_a^{-1}\boldsymbol{\Theta}'^{-1}\mathbf{D}_\rho$$

except that, instead of (15),

$$\mu_i = \sum_j L_{ij}\theta_{ij}/p_j - (\sum_j L_{ij}\theta_{ij})/\phi^* + K_i(1/\phi^* - 1).$$

We can now note the interesting fact that the factor K_i has barely any effect on the efficiency of the $\check{\rho}_j$ and what little effect there is can be shown to be due to dependent mortality (or emigration) and not to dependent movement. Dependent catching, on the other hand, increases the largest terms in μ_i in the ratios L_{ij}:1 and therefore also increases $\boldsymbol{\Sigma}$ approximately in the ratio $\bar{L}$:1 where $\bar{L}$ is a suitably defined average of the L_{ij}.

When $s > t$, instead of (15′),

$$\mu_i = \sum_j L_{ij}\theta_{ij}/p_j - (\sum_j L_{ij}\theta_{ij})/\overline{\phi}^* + K_i(\phi_i/\overline{\phi}^* - \phi^2_i)$$

and similar conclusions obtain.

We have, of course, only dealt with the leading terms in the expansion of $\boldsymbol{\Sigma}$ but the others may still be neglected unless either form of dependence is very considerable.

5. Validity of the Peterson estimate

5·1. It was taken for granted in §§ 2, 3, 4 that the experimenter knows how the population is stratified and is able to conduct the experiment in the requisite manner. Suppose, however, that he does not know or that he is unable to mark the animals in s distinct ways and record

t separate catches. He will then use the Peterson estimate $\tilde{n}_P = ab/c$ (or its unbiased version $(a+1)\,b/(c+1)$).

In § 5·2 we examine the legitimacy of doing this and then, in § 5·3, the appropriateness of the corresponding formula for var $(\tilde{n}_P)$.

The Peterson estimate not only has simplicity to recommend it because even if a stratified experiment is performed, producing an estimate $\tilde{n}$, $\tilde{n}_P$ is preferable if it is valid as it generally has a smaller variance than $\tilde{n}$ (see § 5·5). In § 5·4, therefore, we consider tests for its validity. These tests are also necessary to complement the estimation theory of §§ 2, 3 and they reveal important facts about the movement and catching of the marked individuals.

Except where otherwise stated, there will be no restriction on the relative values of s and t nor on the relative values of the ϕ_i^*. We shall work with the parameters

$$\theta_{ij} = \theta_{ij}^*/\overline{\phi}^*, \quad \phi_i = \sum_j \theta_{ij}, \quad p_j = \overline{\phi}^* p_j^*, \quad n_j = n_j^*/\overline{\phi}^*, \quad n = \sum_j n_j,$$

where $\overline{\phi}^* = (\sum_{i,j} \theta_{ij}^*)/s$.

5·2. Let $\beta = E[b] = \sum_j n_j p_j$ and $\gamma = E[c] = \sum_{i,j} a_i \theta_{ij} p_j$ and let

$$n_P = a\beta/\gamma.$$

For the limit process: $a_i \to \infty$ and $n_j \to \infty$ in such a way that a_i/a, n_j/n and a/n are all constant, n_P/n remains constant and $\tilde{n}_P/n$ converges in probability to n_P/n. Therefore, $\tilde{n}_P$ estimates n if $n_P = n$, that is if

$$n \sum_{i,j} a_i \theta_{ij} p_j = a \sum_j n_j p_j. \tag{25}$$

(25) has an infinity of 'accidental' solutions but we shall be concerned only with those having a simple physical interpretation. One of them is

$$\sum_i a_i \theta_{ij} = a n_j / n, \tag{26}$$

that is, the expected number of marked animals in the jth stratum is proportional to the number of unmarked. A special case of (26) is

$$\theta_{ij} = n_j/n.$$

Next, consider the condition that

$$p_j = p \quad \text{say.} \tag{27}$$

This makes $n_P = n/\sum_i (a_i/a)\,\phi_i = n^*/\sum_i (a_i/a)\,\phi_i^*$ which is equal to $n = n^*/\sum_i (\phi_i^*/s)$ if the ϕ_i^* or the a_i are equal and differs very little if they are not.

A further condition for the validity of $\tilde{n}_P$ can be obtained by making a minor assumption about the relative values of $\mathbf{n}$ and $\boldsymbol{\Theta}$. It is that there exists $\boldsymbol{\xi} = (\xi_i)$ such that

$$n_j = \sum_i \xi_i \theta_{ij}. \tag{28}$$

$\boldsymbol{\xi}$ always exists if the movements of the marked and unmarked animals are the same, since it may then be taken equal to $\mathbf{m}$, the vector of first-sample strata sizes. Otherwise, if $s \geqslant t$, $\boldsymbol{\xi}$ exists provided $\boldsymbol{\Theta}$ is of full rank t. (When $s = t$, $\boldsymbol{\xi}' = \boldsymbol{\eta}' = \mathbf{n}'\boldsymbol{\Theta}^{-1}$ and when $s > t$ there is an infinity of such $\boldsymbol{\xi}$, some of which at least are in the form $\boldsymbol{\eta}' = \mathbf{n}'(\mathbf{U}\boldsymbol{\Theta})^{-1}\,\mathbf{U}$ where $\mathbf{U}$ is any

matrix of the type described in § 2·6.) If $s < t$, $\boldsymbol{\xi}$ does not necessarily exist. Anyhow, assuming that there is a $\boldsymbol{\xi}$ satisfying (28), (25) becomes

$$n \sum_{i,j} a_i \theta_{ij} p_j = a \sum_{i,j} \xi_i \theta_{ij} p_j. \tag{29}$$

Consider

$$\sum_j \theta_{ij} p_j = \overline{p} \quad \text{say.} \tag{30}$$

This makes

$$n_P = \sum_i \xi_i = n^* / \sum_i (\xi_i / \sum_j \xi_j)\, \phi_i^*$$

which equals n if the ϕ_i^* or ξ_i are equal and, otherwise, will not usually differ very much from n. If $\sum_j \theta_{ij} = 1$, (27) is a special case of (30). Another special case of (30) is

$$\theta_{ij} = \theta_j \quad \text{say.}$$

Lastly, when the movement patterns are the same, (26) becomes $\sum_i a_i \theta_{ij} = (a/n) \sum_i m_i \theta_{ij}$ and a particular solution of this equation is

$$a_i / m_i = a/n,$$

the first-sample counterpart of $p_j = p$.

5·3. Having listed the various conditions under which $\tilde{n}_P$ is valid, namely (26), (27) and (30), let us suppose that the experimenter assumes as a matter of faith that one of them holds true and conducts an unstratified experiment. We now check that the formula for var $(\tilde{n}_P)$ that he is obliged to use is an appropriate one.

It is an easy matter to show that

$$E[(\tilde{n}_P - n_P)^2] \sim (\beta^2 a^2/\gamma^4) \operatorname{var}(c) + (a^2/\gamma^2) \operatorname{var}(b) \tag{31}$$

by retaining only leading terms as in §§ 3·2, 3·3. Now

$$\operatorname{var}(c) = \sum_i \operatorname{var}(c_{i\cdot}) = \sum_i a_i (\sum_j \theta_{ij} p_j)(1 - \sum_j \theta_{ij} p_j)$$

$$\leqslant (\sum_{i,j} a_i \theta_{ij} p_j)(1 - \sum_{i,j} a_i \theta_{ij} p_j / a) = a(\gamma/a)(1 - \gamma/a) = \text{'var}(c)\text{'} \quad \text{say.}$$

$\operatorname{var}(c) = $ 'var (c)' only when $\sum_j \theta_{ij} p_j = \overline{p}$ say, that is when c is a true binomial variable. The experimenter has no other course but to use 'var (c)' in (31) and we see that, in doing so, he overestimates var (c) if anything and therefore errs on the right side; but the difference is extremely small. Similarly, 'var (b)' $= n(\beta/n)(1 - \beta/n\overline{\phi}^*)$ overestimates

$$\operatorname{var}(b) = \sum_j n_j p_j (1 - p_j/\overline{\phi}^*)$$

slightly unless $p_j = p$ say. Inserting 'var (c)' and 'var (b)' in (31),

$$\text{'}E[(\tilde{n}_P - n_P)^2]\text{'} \sim (\beta^2 a^2/\gamma^3)(1 - \gamma/a) + (a^2\beta/\gamma^2)(1 - \beta/n\overline{\phi}^*) \tag{32}$$

which is estimated by replacing β by b, γ by c and n by $\tilde{n}_P$.

If there is movement dependence among the marked animals, it has a small but negligible effect on the second of the four terms in (32) and, in view of the fact that the first term is by far the largest, we can certainly ignore this effect. Catch dependence does increase the largest term, however, as in § 4·4. For, if a_ω, c_ω and p_ω denote the number of marked animals present, the number caught and the marginal probability of being caught in a given subregion and, if $E[(c_\omega - a_\omega p_\omega)^2 | a_\omega] = L a_\omega p_\omega (1 - p_\omega)$, the first term in (32) is multiplied by L.

5·4. Reverting to the stratified experiment, we turn from estimation to testing hypotheses. In most cases, the likelihood-ratio method is an obvious one to use and a great variety of hypotheses which might be of interest in particular experiments can be tested in this way. However, our attention will be confined to those tests which bear on the validity of $\tilde{n}_P$ and, moreover, which involve only simple functions of $\{a_i\}$, $\{b_j\}$, $\{c_{ij}\}$. Consider

$$H_1\colon \sum_j \theta_{ij} p_j = \overline{p}, \text{ say}; \quad H_2\colon p_j = p \text{ say}; \quad H_3\colon \sum_i a_i \theta_{ij} = a n_j/n;$$

$$H_4\colon \theta_{ij} = \theta_j \text{ say}; \quad H_5\colon \theta_{ij} = n_j/n.$$

Let H denote the general hypothesis which puts no restrictions on $\{\theta_{ij}\}$, $\{p_j\}$ except, of course, $\sum_{i,j} \theta_{ij} = s$.

On H, the $\psi_{ij} = \theta_{ij} p_j$ are independent and, substituting $\tilde{\psi}_{ij} = c_{ij}/a_j$, the maximum value of the log-likelihood of $\{\theta_{ij}\}$, $\{p_j\}$ is

$$\tilde{L} = \sum_{i,j} c_{ij} \log c_{ij} + \sum_i (a_i - c_{i.}) \log (a_i - c_{i.}) - \sum_i a_i \log a_i.$$

On H_1, the log-likelihood is

$$L_1 = \sum_{i,j} c_{ij} \log \kappa_{ij} + c \log \overline{p} + (a-c) \log (1-\overline{p}),$$

where $\kappa_{ij} = \theta_{ij} p_j / \overline{p}$. Maximizing subject to $\sum_j \kappa_{ij} = 1$,

$$\tilde{L}_1 = \sum_{i,j} c_{ij} \log c_{ij} - \sum_i c_{i.} \log c_{i.} + c \log c + (a-c) \log (a-c) - a \log a.$$

The number of independent, identifiable parameters in L is $N(\psi) = st$ and the number in L_1 is $st - s + 1$. To test H_1 against H we use the fact that, on H_1, $2(\tilde{L} - \tilde{L}_1) = \chi^2_{s-1}$ approximately. Now

$$2(\tilde{L} - \tilde{L}_1) = 2 \sum_i c_{i.} \log (a c_{i.}/a_i c) + 2 \sum_i (a_i - c_{i.}) \log (a(a_i - c_{i.})/(a-c) a_i)$$

and this expression is asymptotically equivalent to

$$\sum_i \frac{(c_{i.} - a_i c/a)^2}{a_i c/a} + \sum_i \frac{(a_i - c_{i.} - a_i(a-c)/a)^2}{a_i (a-c)/a}.$$

Therefore, the test is asymptotically equivalent to a χ^2 goodness-of-fit test for proportionality in the last two columns of (33).

$$\begin{array}{ccc|c|c} c_{11} & \cdots & c_{1t} & a_1 - c_{1.} & c_{1.} \\ \vdots & & \vdots & \vdots & \vdots \\ c_{s1} & \cdots & c_{st} & a_s - c_{s.} & c_{s.} \\ \hline c_{.1} & \cdots & c_{.t} & & \\ \hline b_1 & \cdots & b_t & & \end{array} \tag{33}$$

This is hardly surprising when it is remembered that, on H_1, $E[c_{i.}] = a_i \overline{p}$.

There is no very satisfactory test of H_2 and, to see why, three possibilities have to be considered. First, suppose that nothing is assumed about the relative values of the ϕ_i. The $\psi_{ij} = \theta_{ij} p$ are then mathematically independent and L_2 is not identifiably distinct from L. Therefore H_2 cannot be tested. Secondly, if q constraints are imposed on the ϕ_i, where

$0 < q < s-1$, the maximization of L_2 is awkward and, while a likelihood ratio test based on χ^2_q can be constructed, it does not take a simple form. Lastly, if it is assumed that

$$\phi_i = \sum_j \theta_{ij} = 1,$$

all i,
$$L_2 = \sum_{i,j} c_{ij} \log \theta_{ij} + c \log p + (a-c) \log (1-p),$$

and this is not identifiably distinct from L_1. Therefore, in attempting to test the hypothesis: p_j independent of j and $\sum_j \theta_{ij} = 1$, one is really testing the more general hypothesis: $\sum_j \theta_{ij} p_j$ independent of i.

As H_3 involves the n_j it cannot be tested by the likelihood-ratio method without involving the density $p[\{b_j\}]$ as well as $p[\{c_{ij}\}|\{a_i\}]$. We shall not do this but, instead, simply note that it is equivalent to observe that $E[c_{.j}] = (\sum_i a_i \theta_{ij}) p_j$ and $E[b_j] = n_j p_j$ and that a χ^2_{t-1} test of proportionality in the last two rows of (33) tests the hypothesis

$$\sum_i a_i \theta_{ij} = \nu n_j, \quad \text{or} \quad \sum_i a_i \theta_{ij} = (\sum_i a_i \phi_i)\, n_j/n.$$

This is not quite the same as H_3 and, if true, implies that $n_P = n/\{\sum_i (a_i/a)\, \phi_i\}$ rather than n but, as we pointed out in § 5·2, this sort of difference is immaterial.

H_4 is a particular case of H_1 and would not therefore be tested unless H_1 was accepted first. On H_4, the log-likelihood is $L_4 = \sum_j c_{.j} \log \psi_j + (a-c) \log (1 - \sum_j \psi_j)$, where $\psi_j = \theta_j p_j$. L_4 is maximized when $\psi_j = c_{.j}/a$ giving $\tilde{L}_4 = \sum_j c_{.j} \log c_{.j} + (a-c) \log (a-c) - a \log a$. On H_4, $2(\tilde{L} - \tilde{L}_4) = \chi^2_{st-t}$ approximately and this is equivalent to a χ^2 test on the contingency table formed by the first s rows and $t+1$ columns of (33). The $\chi^2_{(s-1)(t-1)}$ test on the first s rows and t columns, that is on the c_{ij} alone, is easily seen to test H_4 against H_1 or, viewed differently, it tests the hypothesis: $\theta_{ij} = \phi_i \theta_j$ against H.

H_5 cannot be tested as L_5 is not identifiably distinct from L_4.

5·5. The theoretical discussion of the stratified experiment is concluded by a comparison of the variances of $\tilde{n}$ and $\tilde{n}_P$. Much simplicity is gained and nothing essential is lost by assuming in what follows that all $\phi_i = 1$. One advantage of this assumption is that it makes $n_P = n$ exactly whenever $\tilde{n}_P$ is valid. Writing $n_P = n$, the formulae to be compared are

$$E[(\tilde{n}_P - n)^2] \sim (\beta^2 a^2/\gamma^4) \operatorname{var}(c) + (a^2/\gamma^2) \operatorname{var}(b) = P_1 + P_2 \text{ say},$$

and
$$E[(\tilde{n} - n)^2] \sim \sum_i \eta_i^2 \mu_i / a_i + \sum_j n_j (1/p_j - \overline{\phi}^{*-1}) = S_1 + S_2 \text{ say},$$

where $\boldsymbol{\eta}' = \mathbf{n}'\boldsymbol{\Theta}$ or $\mathbf{n}'(\mathbf{U}\boldsymbol{\Theta})^{-1}\mathbf{U}$ according as $s = t$ or $s > t$, and $\mu_i = \sum_j \theta_{ij}/p_j - 1$.

It was observed in § 5·3 that $P_2 = (a^2/\gamma^2) \sum_j n_j p_j (1 - p_j/\overline{\phi}^*) \leqslant (a^2/\gamma^2) \beta (1 - \beta/n\overline{\phi}^*)$ with equality only if $p_j = p$. Therefore, since $a\beta/\gamma = n$, $P_2 \leqslant n^2/\beta - n/\overline{\phi}^*$. In much the same way, $S_2 \geqslant n^2/\beta - n/\overline{\phi}^*$ with equality only if $p_j = p$. Therefore $S_2 \geqslant P_2$.

We also observed in § 5·3 that $P_1 \leqslant (\beta^2 a^2/\gamma^4) \gamma (1 - \gamma/a) = n^2/\gamma - n^2/a$ with equality only if $\sum_j \theta_{ij} p_j = \overline{p}$. To compare S_1 with P_1, it is necessary to consider separately the conditions under which $\tilde{n}_P$ is valid and these boil down to either H_1 or H_3.

On H_1, $$\sum_j \theta_{ij} p_j = \bar{p} \quad \text{or} \quad \sum_j \theta_{ij}(p_j - \bar{p}) = 0. \tag{34}$$

Now, for $\tilde{n}$ to be consistent it is necessary that $\boldsymbol{\Theta}$ is of full rank t and, this being so, (34) implies $p_j = \bar{p}$. Thus, it is sufficient to consider H_2. On H_2, $\mu_i = 1/p - 1 = a/\gamma - 1$. Therefore

$$S_1 = (a/\gamma - 1)\sum_i \eta_i^2/a_i = (a/\gamma - 1)\,[n^2/a + \sum_{i<l} (a_i\eta_l - a_l\eta_i)^2/a_i a_l a]$$

$$\geqslant (a/\gamma - 1)\, n^2/a = n^2/\gamma - n^2/a \geqslant P_1.$$

On H_3, $$\sum_i a_i \theta_{ij} = a n_j/n.$$

Also, $$\sum_i \eta_i \theta_{ij} = n_j. \tag{35}$$

Therefore

$$S_1 = \sum_i (\eta_i/a_i)^2 a_i \mu_i = (n^2/a^2)\sum_i a_i\mu_i + 2(n/a)\sum_i (\eta_i/a_i - n/a)\, a_i\mu_i + \sum_i (\eta_i/a_i - n/a)^2 a_i \mu_i.$$

The first of these terms is $(n^2/a^2)\,[\sum_i a_i\theta_{ij}/p_j - a]$ and, using H_3, equals

$$(n/a)\,[\sum_j n_j/p_j - n] = (n/a)\, S_2 \geqslant n^2/a\beta - n^2/a = n^2/\gamma - n^2/a.$$

Using H_3 and (35), the second term is zero. The third term is greater than or equal to zero. (It equals zero when $s = t$ since H_3 and (35) imply that $a_i = (a/n)\,\eta_i$. When $s > t$, however, $\eta_i/a_i \neq n/a$ in general.) Therefore, $S_1 \geqslant n^2/\gamma - n^2/a \geqslant P_1$.

Thus, in all cases when $\tilde{n}_P$ is valid,

$$E[(\tilde{n} - n)^2] \sim S_1 + S_2 \geqslant P_1 + P_2 \sim E[(\tilde{n}_P - n)^2]$$

with equality only if $p_j = p$ and $\eta_i/a_i = n/a$. The actual difference $S_1 + S_2 - P_1 - P_2$ is easily found from the above and, in practice, can be quite substantial.

6. Analysis of sockeye salmon data

6·1. In an experiment reported by Schaeffer (1951), both stratifications were with respect to time instead of place. The population comprised all adult sockeye salmon who passed a certain point of a river during a period of $s = 8$ weeks on their way up-stream to their spawning grounds. The fish were sampled and tagged according to the week in which they passed this point. Provided they succeed in reaching the spawning grounds, most adult salmon die after spawning. In this case, the deaths took place over a period of $t = 9$ weeks, and, during each of these weeks, a number of dead fish were recovered, presumably very soon after death. As Schaeffer's paper is not easily obtainable for reference, the data of his experiment are reproduced in Table 1. The frequencies in some of the outer weeks are too small to be used in what is essentially a large-sample theory and we have therefore reduced both s and t to four by grouping the first 3 and last 3 weeks of tagging into single strata and the first 3 and last 4 weeks of recovery into single strata. The new values of $\{a_i\}$, $\{b_j\}$, $\{c_{ij}\}$ are given in Table 2.

$\{n_j^*\}$ and $\{p_j^*\}$ have the usual interpretations but θ_{ij}^* now signifies the probability of dying in the jth stratum for a fish tagged in the ith and ϕ_i^* the probability of dying on the spawning grounds during the 9 weeks period. $1 - \phi_i^*$ therefore represents the probability of dying before reaching the spawning grounds or of surviving until after this period. (A small

percentage of salmon do manage to reach the sea alive and return to spawn again. See Jones (1959).)

It is soon apparent that the Peterson estimate is not valid. For, in testing H_1 which specified proportionality of the vectors $(c_{i.})$, $(a_i - c_{i.})$, we obtain $\chi^2_3 = 16{\cdot}91$ and the 0·1 % value is 16·27. The vectors $(c_{.j})$, (b_j) are so obviously not proportional that there is no need to apply a χ^2 test to see that H_3 is unacceptable.

Table 1. *Schaeffer's data.* $\{c_{ij}\}$ *with* $\{a_i\}$, $\{b_j\}$; $s = 8$, $t = 9$

Week of tagging (i)	Week of recovery (j)									Totals	a_i
	1	2	3	4	5	6	7	8	9		
1	1	.	2	.	.	.	.	.	.	3	15
2	1	3	7	.	.	.	.	.	.	11	59
3	1	11	33	24	5	1	.	1	.	76	410
4	.	5	29	79	52	3	2	7	3	180	695
5	.	.	11	67	77	2	16	7	3	183	773
6	.	.	.	14	25	3	10	6	2	60	335
7	.	.	.	.	.	.	1	5	.	6	59
8	.	.	.	.	.	.	1	.	.	1	5
Totals	3	19	82	184	159	9	30	26	8	520	2351
b_j	16	113	718	2664	3317	635	1217	904	368	9952	

Table 2. *Schaeffer's data.* $\{c_{ij}\}$ *with* $\{a_i\}$, $\{b_j\}$; $s = 4$, $t = 4$

	c_{i1}	c_{i2}	c_{i3}	c_{i4}	$c_{i.}$	$a_i - c_{i.}$	a_i	$c_{i.}/a_i$
c_{1j}	59	24	5	2	90	394	484	0·186
c_{2j}	34	79	52	15	180	515	695	0·259
c_{3j}	11	67	77	28	183	590	773	0·237
c_{4j}	0	14	25	28	67	332	399	0·168
$c_{.j}$	104	184	159	73				
b_j	847	2664	3317	3124				

6·2. Applying the theory of § 2·5, it is found, on evaluating $\tilde{\mathbf{p}} = \mathbf{C}^{-1}\mathbf{a}$, that

$$\tilde{p}_1 = 0{\cdot}1318, \quad \tilde{p}_2 = 1{\cdot}9461, \quad \tilde{p}_3 = 0{\cdot}1947, \quad \tilde{p}_4 = 0{\cdot}1063.$$

The unsatisfactory value of $\tilde{p}_2$ may be just a symptom of the general inaccuracy of capture-recapture estimation or it may indicate that the model is incorrect in assuming that the ϕ_i^* are equal, this being necessary for the consistency of $\tilde{p}_2$. Both of these explanations are probably correct but while nothing can be done about the first, we can act on the second. The $c_{i.}/a_i$ indicate where the possible differences in the ϕ_i^* lie, the middle two being appreciably larger than the outer two. Let us therefore estimate subject only to the two constraints: $\phi_1^* = \phi_4^*$, $\phi_2^* = \phi_3^*$. This necessitates a reduction of t from four to three and, consequently, a grouping of two of the second-sample strata. It is permissible to group the jth and kth strata if (i) $(\theta_{ij}p_j + \theta_{ik}p_k)/(\theta_{ij} + \theta_{ik})$ is independent of i, in particular if (ii) $p_j = p_k$ or (iii) θ_{ij}/θ_{ik} is independent of i. (ii) cannot be tested but (iii) can as it implies proportionality of the jth and kth columns of (c_{ij}). In this case, the columns which are nearest to being proportional are the third and fourth and we therefore group them. (The hypothesis of proportionality is rejected at the 0·1 % level but, even so, if p_3 is not too different from p_4, (i) will hold approximately true.)

6·3. Applying the theory of §2·4 to

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0{\cdot}25 & 0{\cdot}25 & 0{\cdot}25 & 0{\cdot}25 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 59 & 24 & 7 \\ 34 & 79 & 67 \\ 11 & 14 & 105 \\ 0 & 14 & 53 \end{pmatrix},$$

$$\mathbf{a}' = (484, 695, 773, 399), \quad \mathbf{b}' = (847, 2664, 6441)$$

and evaluating $\tilde{\boldsymbol{\rho}} = (\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C})^{-1}\mathbf{v}$, we obtain

$$\tilde{\boldsymbol{\rho}}' = (6{\cdot}021, 1{\cdot}607, 6{\cdot}397) \quad \text{or} \quad \tilde{p}' = (0{\cdot}1661, 0{\cdot}6223, 0{\cdot}1563).$$

$\tilde{p}_2$ now lies in [0,1] but is still curiously high. The other estimates are

$$\tilde{\boldsymbol{\theta}} = \mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C}\mathbf{D}_{\boldsymbol{\rho}} = \begin{pmatrix} 0{\cdot}7339 & 0{\cdot}0797 & 0{\cdot}0925 \\ 0{\cdot}2945 & 0{\cdot}1827 & 0{\cdot}6167 \\ 0{\cdot}0857 & 0{\cdot}1393 & 0{\cdot}8689 \\ 0{\cdot}0000 & 0{\cdot}0564 & 0{\cdot}8497 \end{pmatrix}$$

and, summing the rows,

$$\tilde{\phi}_1 = \tilde{\phi}_4 = 0{\cdot}9061, \quad \tilde{\phi}_2 = \tilde{\phi}_3 = 1{\cdot}0939.$$

Also $$\tilde{\mathbf{n}}' = \tilde{\boldsymbol{\rho}}'\mathbf{D}_{\mathbf{b}} = (5{,}099,\ 4{,}282,\ 41{,}204),$$

and $$\tilde{n} = 50{,}585.$$

The estimated variance-covariance matrix of $\tilde{\boldsymbol{\rho}}$ is

$$\tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} 9{\cdot}96 & -14{\cdot}84 & 6{\cdot}31 \\ -14{\cdot}84 & 23{\cdot}58 & -10{\cdot}32 \\ 6{\cdot}31 & -10{\cdot}32 & 4{\cdot}78 \end{pmatrix}.$$

Note the very high variance of $\tilde{\rho}_2$. It is unlikely that there was any catch dependence in this experiment so that we need make no mental reservations about $\tilde{\boldsymbol{\Sigma}}$ underestimating $\boldsymbol{\Sigma}$. Next,

$$E[(\tilde{\mathbf{n}} - \mathbf{n})(\tilde{\mathbf{n}} - \mathbf{n})'] = 10^6 \begin{pmatrix} 7{\cdot}168 & -33{\cdot}474 & 34{\cdot}441 \\ -33{\cdot}474 & 167{\cdot}347 & -177{\cdot}113 \\ 34{\cdot}441 & -177{\cdot}113 & 198{\cdot}694 \end{pmatrix}$$

and $$E[(\tilde{n} - n)^2] = 20{\cdot}916 \times 10^6.$$

It is noteworthy that var $(\tilde{n}_2)$ and var $(\tilde{n}_3)$ are each considerably larger than var $(\tilde{n})$.

Although $\tilde{n}_P$ is invalid, it is worth evaluating it and its variance to compare with $\tilde{n}$ and var $(\tilde{n})$. Using the unbiased version $(a+1)\,b/(c+1)$,

$$\tilde{n}_P = 44{,}927,$$

and $$E[(\tilde{n}_P - n_P)^2] = 3{\cdot}181 \times 10^6.$$

The latter is a good deal smaller than var $(\tilde{n})$ as might be expected since the $\tilde{p}_j$ differ and, even more so, the $\tilde{\eta}_i/a_i$. For

$$\tilde{\boldsymbol{\eta}}' = \mathbf{b}'(\mathbf{U}\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C})^{-1}\,\mathbf{U} = (9896,\ -20{,}728,\ 46{,}020,\ 15{,}397).$$

6·4. Finally, we consider how far we were justified in §6·2 in inferring that the ϕ_i differ. It can be shown that

$$\text{var}\,(\tilde{\phi}_1) \sim \mu_1/a_1 + \boldsymbol{\theta}_1'(\mathbf{U}\boldsymbol{\Theta})^{-1}\,\mathbf{U}\mathbf{D}_{\boldsymbol{\mu}}\mathbf{D}_{\mathbf{a}}^{-1}\,\mathbf{U}'(\mathbf{U}\boldsymbol{\Theta})'^{-1}\,\boldsymbol{\theta}_1 - 2\mu_1/a_1\,\boldsymbol{\theta}_1'(\mathbf{U}\boldsymbol{\Theta})^{-1}\,\mathbf{U}\boldsymbol{\delta}_1, \qquad (36)$$

where $\boldsymbol{\theta}_1' = (\theta_{1j})$ and $\boldsymbol{\delta}_1 = (\delta_{i1})$. Using this formula, we find that the estimated standard error of $\hat{\phi}_1$ is 0·1509, and therefore the difference between $\hat{\phi}_1 = 0{\cdot}9061$ and 1 is not significant. However, this non-significance may probably be attributed mainly to the insensitivity of the estimation. The variances of $\hat{\phi}_2$, $\hat{\phi}_3$, $\hat{\phi}_4$ can be found by formulae similar to (36) and a useful check on the computation is provided by verifying that they are all the same.

I wish to thank Professors M. S. Bartlett and D. V. Lindley for their very helpful comments.

REFERENCE

BEVERTON, R. J. H. & HOLT, S. J. (1957). *On the Dynamics of Exploited Fish Populations.* London: Her Majesty's Stationery Office.

CHAPMAN, D. G. & JUNGE, C. O. (1956). The estimation of the size of a stratified animal population. *Ann. Math. Statist.* **27**, 375–89.

DARROCH, J. N. (1958). The multiple-recapture census. I. Estimation of a closed population. *Biometrika*, **45**, 343–59.

DARROCH, J. N. (1959). The multiple-recapture census. II. Estimation when there is immigration or death. *Biometrika*, **46**, 336–51.

JONES, J. W. (1959). *The Salmon.* London: Collins.

SCHAEFFER, M. B. (1951). Estimation of the size of animal populations by marking experiments. *U.S. Fish and Wildlife Service Fisheries Bulletin*, **69**, 191–203.

Biometrika (1961), **48**, 3 *and* 4, *p.* 261
With 1 *text-figure*
Printed in Great Britain

The fiducial method and invariance

BY D. A. S. FRASER
University of Toronto†

1. INTRODUCTION

The fiducial method for estimating the value of a parameter was introduced by Fisher in 1930. In papers since that time Fisher has frequently discussed aspects of the method and may have, in the view of some readers, modified or altered his ideas concerning the underlying principles and the more general aspects of the method. The method has been frequently criticized adversely in the literature, particularly in recent years. Some of the grounds for this criticism are: conflict with the confidence method in certain problems; non-uniqueness in certain problems; disagreement with 'repeated sampling' frequency interpretations; and a lack of a seemingly-proper relationship with *a priori* distributions.

In his recent book, *Statistical Methods and Scientific Inference*, Fisher (1956) has devoted considerable space to the fiducial method. He states that an essential ingredient for its use is the absence of prior information concerning the value of the parameter being estimated; in his words, 'it is essential to introduce the absence of knowledge *a priori* as a distinctive datum in order to demonstrate completely the applicability of the fiducial method of reasoning to the particular real and experimental cases for which it was developed.' An interpretation of one aspect of this requirement might be that all parameter values are equivalent in the way in which the frequency distribution of the observable variable is related to the parameter value determining that distribution. In § 5 this interpretation is formalized and shown to imply a mathematical model in which the parameter is related to a group of transformations on the sample space.

In §§ 2 and 3 the mathematical model involving transformations is presented on its own merits and in §§ 4 and 5 the fiducial argument for it is developed. A consequence for this model is that the information about the parameter from an observed value of the variable is in the form of a frequency distribution, the fiducial distribution, having a frequency interpretation in terms of a well-defined kind of repeated sampling. This is in agreement with Fisher's statement—'the fiducial argument uses the observations (only) to change the logical status of the parameter from one in which nothing is known of it, and no probability statement can be made, to the status of a random variable having a well-defined distribution.'

Another consequence in the special framework is that there is no need to require the absence of an *a priori* distribution for the parameter. For, if the fiducial distribution is combined in a logical manner with the *a priori* distribution the result is the *a posteriori* distribution of a Bayesian argument—a reassuring result. This is demonstrated in § 9.

A further consequence concerns prior information that the parameter value is restricted to some specified range. This restriction can be used to condition the fiducial distribution yielding a *conditioned* fiducial distribution. A probability combination of such restrictions can in effect generate an *a priori* distribution and an appropriate combination of conditioned fiducial distributions yields the Bayesian *a posteriori* distribution. This is discussed in § 10.

† Present address: Department of Statistics, Stanford University, Stanford, California.

In § 11 consideration is given to the case of two separate sets of variables each producing a fiducial distribution concerning a particular parameter. A logical method is proposed for combining such fiducial distributions to yield a resultant fiducial distribution. The same resultant distribution is obtained when the fiducial distribution from one set is used as a prior distribution for a Bayesian argument on the second. The same distribution is also obtained when the combined system yields a sufficient statistic and thereby a fiducial distribution. This equivalence of the results from three quite different approaches is a reassuring feature for the use of the fiducial method in the transformation framework proposed here.

David Brillinger in an unpublished paper prepared at Princeton University has proved that if a fiducial distribution is used to produce an invariant interval or region, then that interval or region is also of confidence type. The interesting cases are, however, those in which the interval or region is chosen relevant to the particular observed values of the variables and hence in general not of invariant form. The usual fiducial solutions for the Behrens–Fisher problem and the Creasy–Fieller problem (specified variances and covariances) are of this form and for these there is a legitimate frequency interpretation. It is suggested then that the usual criticisms of the Behrens–Fisher solution are made within a certain mathematical framework and that the fault may well lie in the use of that framework to evaluate fiducial probability. It is further suggested that the fiducial method for problems having the transformation form is not merely better but is the only logically justifiable method on the grounds that the scientific method requires the use of all information available.

The frequency interpretation of the fiducial distribution has the form of a conditional distribution of possible values for the parameter given the observed values of the variables. In this form all the familiar methods for handling conditional distributions are available.

2. The transformation model

The term specification is sometimes used to refer to the basic ingredients of a statistical problem—the sample space, the parameter space, and the class of probability distributions. In this section a general type of specification based on transformations is presented and illustrated by a simple example, sampling from a normal distribution. In a later section this type of specification will be related to the requirements that Fisher puts forward for the application of the fiducial argument.

Consider a basic sample space on which there are probability distributions indexed by a parameter θ which takes values in a parameter space Ω. Suppose this specification admits a sufficient statistic x taking values in a derived sample space, $\mathscr{X}$. The term *sufficient statistic* is used here in the Fisherian sense:† the conditional distribution given the statistic does not depend on the parameter θ; no reduction can be made on the statistic (no non-trivial function of the statistic can be taken) without losing the preceding property; the dimension of the statistic is the same as the dimension of the parameter. A statistic possessing only the first two properties is termed *exhaustive* by Fisher (the term minimal sufficient is often used in more mathematically oriented papers). Further assumptions in this section will, in the present context, make precise the term dimension.

† As formalized from papers such as Fisher (1934, 1948, 1950).

Suppose now that there is a *group* $\mathscr{G}$ of transformations on the sample space that is meaningful or reasonable in terms of the physical problem from which the statistical problem here considered might have been abstracted. A class $\mathscr{G}$ of transformations is a group of transformations if the following conditions are fulfilled: if g and h are transformations belonging to $\mathscr{G}$, then the composite transformation hg obtained by first applying g and then applying h belongs to $\mathscr{G}$; if g belongs to $\mathscr{G}$, then the inverse transformation g^{-1} exists and belongs to $\mathscr{G}$.

Also suppose that these transformations relate naturally to the sufficient statistic and to the parameter; precisely, suppose the following three properties hold. Firstly, the transformations on the basic sample space induce transformations on the values of the sufficient statistic: if $g \in \mathscr{G}$ and $x \in \mathscr{X}$, then the transformation g applied to basic sample points that yield a common value x for the sufficient statistic will carry them into basic sample points all of which yield a value gx for the sufficient statistic. Thus the transformations can be viewed as applying to the space $\mathscr{X}$. Secondly, there is one and only one transformation carrying any point x in $\mathscr{X}$ into any other point x' in $\mathscr{X}$. Then, by relating some arbitrary reference point, say, x_0 in $\mathscr{X}$, to the identity element e in $\mathscr{G}$, an arbitrary point x in $\mathscr{X}$ can be described by means of the unique transformation that produces x from the reference point x_0. Thus in a sense $\mathscr{X}$ and $\mathscr{G}$ are isomorphic. Thirdly, a transformation g carries a variable x with a θ distribution into a variable gx having a distribution $g^*\theta$ in Ω (that is, within the specification) and there is one and only one transformation such that any point θ in Ω is carried into any other point θ' in Ω. Then, by using a reference point θ_0 in Ω, a general parameter value θ can be described by the unique transformation that produces it from the reference point θ_0. Thus in a sense $\mathscr{X}$, $\mathscr{G}$, and Ω are isomorphic one to another.

The simple example of sampling from a normal distribution illustrates the above ideas quite well. Let $(x_1, \ldots, x_n)$ be a sample from a normal distribution with mean μ and standard deviation σ with no restrictions on the ranges of μ and σ. The parameter is $\theta = (\mu, \sigma)$ and the parameter space is $\Omega = (-\infty, \infty) \times (0, \infty)$, the upper half plane. The combination of the sample mean $\bar{x}$ and the sample standard deviation s is a sufficient statistic in the Fisherian sense. Thus the general x is here equal to $(\bar{x}, s)$ and the sample space $\mathscr{X}$ is the upper half plane. Notice the identity of the space $\mathscr{X}$ and Ω.

For many physical problems yielding this particular specification a linear transformation that does not reverse the direction on the axis of measurement has a certain naturalness. In fact, such a transformation corresponds to a change of origin or zero for measurement and a change of scale or unit for measurement. Usually the origin and unit are a matter of convention and in no sense an intrinsic part of the applied problem.

Let $[a, b]$ designate a linear transformation that moves the origin by an amount a and changes the scale by the positive factor b. This transformation applies on the basic sample space as follows

$$[a, b](x_1, \ldots, x_n) = (a + bx_1, \ldots, a + bx_n).$$

The induced transformation on the space of the sufficient statistic is then with simple algebra seen to be

$$[a, b](\bar{x}, s) = (a + b\bar{x}, bs).$$

The class of such transformations

$$\mathscr{G} = \{[a, b] \mid -\infty < a < \infty, 0 < b < \infty\}$$

is a group; this follows from the closure of $\mathscr{G}$ under the formation of products and inverses, the formulas for which are easily seen to be

$$[c, d][a, b] = [c + da, db]; \quad [a, b]^{-1} = [-a/b, 1/b].$$

There is a unique linear transformation carrying a value of the sufficient statistic $(\bar{x}', s')$ into any other value $(\bar{x}'', s'')$. Accordingly, by choice of a reference point, say the particular simple choice (0, 1), the transformations can be used as name tags for the points in the space $\mathscr{X}$, in fact the transformation $[\bar{x}, s]$ carries the simple reference point (0, 1) into the general point $(\bar{x}, s)$ yielding the following correspondence between transformations and sample points

$$[\bar{x}, s] \leftrightarrow (\bar{x}, s).$$

Under a linear transformation $[a, b]$, a sample from a normal distribution (μ, σ) is carried into a sample from the normal distribution $(a + b\mu, b\sigma)$. Thus, in notation suggested by the earlier part of this section,

$$[a, b]^*(\mu, \sigma) = (a + b\mu, b\sigma).$$

There is a unique transformation carrying any parameter value into any other; accordingly, the transformations can be used as name tags for parameter points and, with reference point (0, 1), the correspondence is

$$[\mu, \sigma] \leftrightarrow (\mu, \sigma).$$

Thus the sample space, the transformation space, and the parameter space are isomorphic.

A specification satisfying the transformation properties introduced in this section has the merit that any sample point has a *position relative* to any other sample point (as expressed by the transformation carrying one into the other) and any parameter point has a *position relative* to any other parameter point. The next two sections will establish that a sample point has a *position relative* to a parameter point and, more important for the fiducial approach, a parameter point has a *position relative* to a sample point.

Some discussion of transformations and the invariance associated with them may be found in Lehmann (1959), Chapter 6.

3. A MORE GENERAL MODEL

In problems having merely an exhaustive statistic a natural group of transformations may exist and yield the essentials, for present purposes, of the structure in the preceding section.

Let θ in Ω be a parameter indexing the distributions of an exhaustive statistic x in $\mathscr{X}$. Let g in a group $\mathscr{G}$ be a typical transformation on a basic sample space yielding a transformation on values of the exhaustive statistic x. The group $\mathscr{G}$ of transformations g induces a partition of the space $\mathscr{X}$ into invariant sets in the following manner. Let S_x be the set of images of x under the transformations in $\mathscr{G}$

$$S_x = \{x' | x' = gx \text{ for } g \text{ in } \mathscr{G}\}.$$

S_x can reasonably be called the *trajectory* or *orbit* of x under the group $\mathscr{G}$. It is easily proved that any S_x contains the point x and that any two S's are either identical or disjoint; the S's thus form a partition of $\mathscr{X}$.

Suppose that there is a unique transformation carrying any point x in a set S into any other point x' in that S. The transformations can then be used as name tags for points within a set S (relevant to some reference point).

Also suppose that a transformation g carries a variable x with a θ distribution into a variable gx with a distribution $g^*\theta$ in Ω and that there is a unique transformation such that any point θ in Ω is carried into any other point θ' in Ω in the manner just described. The transformations can then be used as name tags for points in Ω (relevant to some reference point).

By taking a very general view of a statistic as a function on a sample space, it is possible to view S_x as a statistic, that function that produces a set of the partition from a point x in $\mathscr{X}$. As such, it has a distribution. In fact, here it has a distribution that does not depend on the parameter θ; it has a fixed distribution. Such a statistic with a fixed distribution is called an *ancillary* statistic by Fisher. This property of the statistic S_x is easily demonstrated. For we note that $S_x = S_{gx}$ for all transformations g in $\mathscr{G}$. If x is treated as a variable with a θ distribution, then gx is a variable with a $g^*\theta$ distribution. The identity of S_x and S_{gx} shows that the distribution is the same; in other words that S_x has the same distribution when derived from θ as when derived from $g^*\theta$, and hence the same regardless of the parameter value.

In such a situation involving an ancillary statistic, Fisher suggests that the problem be analyzed conditionally given the ancillary statistic. Some grounds for this may be found in Fisher (1948), in Buehler (1959), and in Wallace (1959). A conditional analysis within a set S and using transformations $\mathscr{G}$ and parameter space Ω reduces the specification to the form considered in the preceding section.

As an example let $(x_1, \ldots, x_n)$ be a sample from a distribution with known form but unknown location parameter μ and unknown scale parameter σ. Suppose the distribution has a density function. The density function then has the form

$$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right),$$

where the function f is specified.

The variable $(x_1, \ldots, x_n)$ can be reduced to the order statistic† $\{x_1, \ldots, x_n\}$ with the conditional distribution being independent of the parameters. However, except for special functions f, no further reduction can be made with the conditional distribution remaining independent of the parameters. Thus, in general, the order statistic $\{x_1, \ldots, x_n\}$ is exhaustive. For purposes of illustration here it suffices and is simpler to work with $(x_1, \ldots, x_n)$ rather than the trivial reduction, the order statistic.

As a natural group of transformations for this problem, consider the group $\mathscr{G}$ of linear transformations discussed in the preceding section. Under the group $\mathscr{G}$ a point $(x_1, \ldots, x_n)$ can be moved parallel to the vector $(1, \ldots, 1)$ and radially in and out from that vector. These 'motions' from an initial point generate a set S of the partition of the sample space. An element $[a, b]$ of the group $\mathscr{G}$ operates in the following manner.

$$[a, b]\,(x_1, \ldots, x_n) = (a + bx_1, \ldots, a + bx_n),$$

$$[a, b]\,(\mu, \sigma) = (a + b\mu, b\sigma).$$

The conditions proposed earlier in this section are satisfied by this example.

The sets S can be easily indexed and so can the points within a set S. Take any location and scale statistics; a natural choice is $(\bar{x}, s)$, the sample mean and standard deviation. The statistic $(\bar{x}, s)$ can be used to describe points within a set S, and on these points the transformations operate as follows

$$[a, b]\,(\bar{x}, s) = (a + b\bar{x}, bs).$$

Take any invariant statistics that complete the description of a sample point; a simple choice is

$$A = \left(\frac{x_1 - \bar{x}}{s}, \ldots, \frac{x_n - \bar{x}}{s}\right)$$

† The *set* of x values without regard to the order in which they appear in the sample point $(x_1, \ldots, x_n)$.

which describes the relative positioning of the sample values with respect to $(\bar{x}, s)$. A 'value' for this statistic determines a set S of the partition. The above statistic is called by Fisher a *configuration statistic*; it is ancillary for this problem.

The approach suggested earlier in this section is to analyse $(\bar{x}, s)$ relevant to (μ, σ) conditionally given the configuration statistic. In other words to contemplate only such other possible samples as have the same configuration as the particular sample observed.

4. A PIVOTAL QUANTITY

Fiducial distributions are usually derived by means of a pivotal quantity. A *pivotal quantity* for a specification is a function of the sufficient statistic and the parameter that has a fixed distribution regardless of the value of the parameter. A simple example is the t-quantity for sampling from a normal distribution, $t = \sqrt{n}(\bar{x}-\mu)/s$; it is a function of the sufficient statistic $(\bar{x}, s)$ and has a fixed distribution, Student's with $n-1$ degrees of freedom. In the presence of an ancillary statistic, the usefulness of a pivotal quantity for fiducial purposes requires that it have a fixed *conditional* distribution given the ancillary statistic.

In the statistical literature there are examples of specifications that have yielded several essentially different pivotal quantities with the result that several fiducial distributions for the parameter have been obtained. Some prominent examples can be found in Mauldon (1955) and Tukey (1957). In this section a pivotal quantity will be developed for the specification described in § 2 and will be shown to be unique in the following sense: any other pivotal quantity invariant in form under the transformations in the group is a function of the pivotal quantity developed here. A consequence of this result is that if the principle is accepted that estimation procedures should be invariant with respect to natural transformations leaving the specification invariant then the fiducial distribution will be uniquely determined. For the Mauldon example there are groups that fulfil the requirements in § 2. If one of these groups is added to the specification as being a natural ingredient of the problem, then with reference to the principle just mentioned the fiducial distribution is uniquely determined.

The development here will be with reference to the model in § 2. The model in § 3 reduces to the simpler model if the sample space is restricted to that given conditionally by the ancillary statistic.

Let x_0, θ_0 be arbitrary but fixed reference points in the sample space $\mathscr{X}$ and the parameter space Ω. Elements of the group $\mathscr{G}$ can then be used to describe sample and parameter points. Let g_x be the unique element of $\mathscr{G}$ that carries x_0 into a general sample point x, and similarly let h_θ be the element that carries θ_0 into a general parameter point θ.

As a transformation on Ω, h_θ carries θ_0 into θ. Therefore, as a transformation on $\mathscr{X}$, h_θ must carry a variable with a θ_0 distribution into a variable with a θ distribution. The inverse h_θ^{-1} then carries a variable with a θ distribution into a variable with a θ_0 distribution.

Consider now $x = g_x x_0$ as a variable with a θ distribution. Applying the transformation h_θ^{-1} produces a variable $h_\theta^{-1} g_x x_0$ with a θ_0 distribution. This variable is a function of θ and x and it has a fixed distribution regardless of the value of the parameter. It is generated by the random variable $h_\theta^{-1} g_x$ treated as a random transformation and applied to the fixed reference point x_0.

Thus $g = h_\theta^{-1} g_x$ is a function of x and θ taking values in $\mathscr{G}$ and having a fixed distribution when x is treated as a variable with a θ distribution; it is a pivotal quantity. *As a function of*

x and θ it is invariant under transformations in $\mathscr{G}$. For, let f be a transformation in $\mathscr{G}$; under f, $x = g_x x_0$ is transformed into $fx = fg_x x_0$, and a distribution $\theta = h_\theta \theta_0$ for x becomes the distribution $f\theta = fh_\theta \theta_0$ for fx. Hence

$$h_\theta^{-1} g_x = (fh_\theta)^{-1} (fg_x).$$

Any function of x and θ that is invariant with respect to $\mathscr{G}$ is now shown to be a function of the pivotal quantity $h_\theta^{-1} g_x$. Let $P(x; \theta)$ be invariant with respect to $\mathscr{G}$: $P(fx; f\theta) \equiv P(x; \theta)$ for f in $\mathscr{G}$. Using this invariance yields

$$\begin{aligned} P(x; \theta) &= P(g_x x_0; h_\theta \theta_0) \\ &= P(h_\theta^{-1} g_x x_0; h_\theta^{-1} h_\theta \theta_0) \\ &= P(h_\theta^{-1} g_x x_0; \theta_0). \end{aligned}$$

Thus $P(x; \theta)$ has been expressed as a function of $h_\theta^{-1} g_x$. From this it follows that any invariant pivotal quantity is a function of the particular pivotal quantity $h_\theta^{-1} g_x$. In its essentials then in terms of its usefulness, the pivotal quantity $h_\theta^{-1} g_x$ is unique—any other pivotal quantity is a reduction from it.

Consider the example involving sampling from a normal distribution. Using $(0, 1)$ as the reference point in both the sample and parameter spaces produces

$$g_x = [\bar{x}, s], \quad h_\theta = [\mu, \sigma].$$

The pivotal quantity then has the form

$$\begin{aligned} g = h_\theta^{-1} g_x &= [\mu, \sigma]^{-1} [\bar{x}, s] \\ &= \left[-\frac{\mu}{\sigma}, \frac{1}{\sigma} \right] [\bar{x}, s] \\ &= \left[\frac{\bar{x} - \mu}{\sigma}, \frac{s}{\sigma} \right]; \end{aligned}$$

it is the essentially unique invariant pivotal quantity. Normal distribution theory shows that its distribution can be described by the pivotal variable

$$g = \left[\frac{z}{\sqrt{n}}, \frac{\chi}{\sqrt{(n-1)}} \right],$$

where z is a standardized normal variable and χ is an independent χ-variable on $n-1$ degrees of freedom.

The distribution of the pivotal variable g can also be obtained by a direct argument using the general theory. g as a variable in $\mathscr{G}$ must carry the fixed point $x_0 = (0, 1)$ into a variable having a $\theta_0 = (0, 1)$ distribution, that is, $g(0, 1) = \{z/\sqrt{n}, \chi/\sqrt{(n-1)}\}$. From this it follows that $g = [z/\sqrt{n}, \chi/\sqrt{(n-1)}]$.

The general discussion earlier in this section showed that the frequency distribution for x produced a fixed frequency distribution for the pivotal variable g. What is perhaps of even greater interest is that *this fixed distribution of the pivotal variable when used in conjunction with the pivotal equation $g = h_\theta^{-1} g_x$ completely describes the specification of the problem.* For, let g be a variable having the fixed pivotal distribution. The equation

$$g = h_\theta^{-1} g_x$$

can be solved yielding

$$\begin{aligned} g_x &= h_\theta g, \\ x = g_x x_0 &= h_\theta g x_0, \end{aligned}$$

which demonstrates that the general frequency distribution for x is obtained by a parameter transformation (h_θ) applied to gx_0; gx_0 is a variable on the sample space $\mathscr{X}$ obtained by a random 'error' transformation g applied to the reference point x_0. The general distribution is thus in a sense merely a relocation of an 'error' distribution about an arbitrary reference point x_0.

5. The fiducial distribution

A fiducial distribution purports to be a frequency distribution of possible values for a parameter as obtained, through frequency information summarized in the specification, from an observed value of a variable. The technique for deriving a fiducial distribution is the following: the observed value of the sufficient statistic is substituted into the pivotal equation; the pivotal variable is treated as a variable with the frequency distribution normally associated with it; the parameter in the equation is treated as a free variable and the distribution of the pivotal variable is transferred to it by the pivotal equation.

Let g be a variable with the fixed pivotal variable distribution; let x be the observed value of the sufficient statistic; and let θ be a variable designating possible values for the parameter in the light of the frequency and observational information. The pivotal equation can be solved† for θ in terms of g and x

$$g = h_\theta^{-1} g_x,$$
$$h_\theta = g_x g^{-1},$$
$$\theta = g_x g^{-1} \theta_0.$$

This last equation gives the fiducial distribution for θ as obtained from the fixed distribution of the pivotal variable.

In an application of this theory the parameter θ has a particular *but unknown* value. This value determines the distribution of the sufficient statistic which in turn produces the observed value x. In such a situation θ is certainly fixed but this does not mean‡ that probability statements cannot be made concerning θ.

In § 4 it was shown that the frequency information available concerning the sufficient statistic relative to the parameter can be summarized in the fixed distribution of the pivotal variable. In terms of repeated sampling from this fixed distribution for the 'error' variable g, there is generated a frequency distribution θ of possible parameter values corresponding to the observed x. This is the fiducial distribution and it has here a frequency interpretation. This interpretation will be elaborated upon in terms of the example involving sampling from a normal distribution; the interpretation, however, applies generally.

Consider again the example. The pivotal variable was obtained in § 4; its distribution can be expressed by

$$g = [z/\sqrt{n}, \chi/\sqrt{(n-1)}].$$

A frequency distribution is generated when this random 'error' transformation is applied to the reference point $(0, 1)$ and the frequency distribution of the observable variable is then obtained by a transformation on the sample space where the transformation is determined by the parameter.

† The preceding section contains the solution for x in terms of g and θ; it describes the specification directly.

‡ Consider a gambling game in which the dice are rolled but concealed from view; the bets made; and the dice exposed. With honesty this game is equivalent to one in which the bets are made before the dice are rolled. Even if a person contemplated only one play I doubt that valid grounds could be given for preferring one to the other.

The formulas earlier in this section produce the fiducial distribution

$$g^{-1}(0,1) = [z/\sqrt{n}, \chi/\sqrt{(n-1)}]\,(0,1)$$
$$= \left(-\frac{z/\sqrt{n}}{\chi/\sqrt{(n-1)}}, \frac{\sqrt{(n-1)}}{\chi}\right),$$
$$(\hat{\mu}, \hat{\sigma}) = g_x g^{-1}(0,1) = [\bar{x}, s]\, g^{-1}(0,1)$$
$$= \left(\bar{x} - \frac{z}{\chi/\sqrt{(n-1)}}\frac{s}{\sqrt{n}}, \frac{\sqrt{(n-1)}}{\chi}s\right).$$

Thus

$$\hat{\sigma} = \frac{\sqrt{(n-1)}}{\chi}s,$$
$$\hat{\mu} = \bar{x} - \frac{z}{\chi/\sqrt{(n-1)}}\frac{s}{\sqrt{n}} = \bar{x} - z\frac{\hat{\sigma}}{\sqrt{n}}$$
$$= \bar{x} - t\frac{s}{\sqrt{n}},$$

where the t-variable implicitly defined is statistically dependent on the χ variable. The fiducial variables $\hat{\mu}, \hat{\sigma}$ are obtained from the observed values $\bar{x}, s$ by the action of the pivotal variables z, χ. It is interesting to note that the fiducial distribution of $\hat{\mu}$ is centred at $\bar{x}$, is scaled by $s/\sqrt{n}$, and has the form of Student's distribution on $n-1$ degrees of freedom.

Suppose a performance of an experiment involving a sample of n from a normal distribution has yielded the observed values $\bar{x}, s$. Now contemplate many past and possible performances of experiments involving samples of n from normal distributions. For most such performances the origin and scaling is purely conventional. To make these other performances comparable to the particular performance in hand, transform each by rescaling and relocating so that the same mean in each case is moved to the value $\bar{x}$ and the standard deviation to the value s. In each case apply conceptually the transformation to the mean of the distribution to yield a value that is suitable for comparison with the values $\bar{x}, s$. The collection of these transformed means generates a frequency distribution and it is fairly easily seen that this frequency distribution is the fiducial t-distribution obtained in the preceding paragraph. From this point of view then the fiducial distribution is a frequency distribution of possible values for the parameter relevant to the particular observed $\bar{x}, s$. It seems quite proper then to use this distribution to make probability statements in which μ, σ appear formally as variables. Such probability statement would be consistent with Fisher's operational definition of probability; see, for example, Fisher (1959).

The interpretation of the fiducial distribution for the general case proceeds in a similar manner. The mathematics associated with it takes the following form. Let x be an observed value of the sufficient statistic. Let X be the observable sufficient statistic with distribution given by

$$X = h_\theta g x_0,$$

where the randomness is produced by the 'error' variable g, h_θ is the parameter transformation, and x_0 is the reference point. θ need not be fixed; but a value for θ and an observed value for g produce an observed value for X.

The transformation that carries X into x is $g_x g^{-1} h_\theta^{-1}$, as is easily checked,

$$g_x g^{-1} h_\theta^{-1} X = g_x g^{-1} h_\theta^{-1} h_\theta g x_0$$
$$= g_x x_0 = x.$$

Applying this transformation to θ produces

$$g_x g^{-1} h_\theta^{-1} \theta = g_x g^{-1} \theta_0$$

which is a variable with distribution generated by the 'error' variable g. This frequency distribution of possible values for θ is just the fiducial distribution that was developed earlier in this section.

With the preceding interpretation for fiducial probability it seems quite proper to refer to it as an *a posteriori* distribution obtained in the absence of an *a priori* distribution.

6. Relationship to Fisher's requirements

Throughout his writings on fiducial inference, Fisher has emphasized that the method be based on an exhaustive statistic, in fact, that it be based either on a sufficient statistic or on an exhaustive statistic that has an ancillary so that with respect to the ancillary the exhaustive statistic is conditionally sufficient. This is the basic assumption for the specifications in §§ 2 and 3.

Near the beginning of § 1 there is a quotation from Fisher describing an 'essential ingredient' for the use of the fiducial method. Following this, there appears my interpretation of one aspect of this ingredient—'that all parameter values are equivalent in the way in which the frequency distribution of the observable variable is related to the parameter value determining that distribution'. Implicitly then, there must be a means for comparing the frequency distribution for one value of θ with the frequency distribution for another value of θ. The only reasonable way of making such comparisons, seemingly, is by means of transformations on the sample space:

Assumption 1. *There is a group $\mathcal{G} = \{g\}$ of* 1-1 *measurable transformations on the sample space $\mathcal{X}$ of the sufficient statistic onto itself.*

That the set of transformations be a group follows from the reasonable requirement that the set of transformations be closed under the formation of products and inverses.

The purpose in having transformations is to permit comparison of various distributions for the sufficient statistic x. A transformation on the sample space must then carry one possible distribution for the sufficient statistic into another possible distribution for it:

Assumption 2. *The class, indexed by θ in Ω, of probability distributions for x is closed under the transformations in $\mathcal{G}$.*

If g is a transformation in $\mathcal{G}$ and x has a distribution determined by θ then the variable gx has a distribution on the sample space $\mathcal{X}$ and by the assumption there is a parameter value for it: $g^*\theta$. The transformation g^* on Ω describes the changes in parameter values corresponding to the changes in distribution determined by g. The transformations $\mathcal{G}^* = \{g^*\}$ form a group and the mapping $g \to g^*$ is a homomorphism; see, for example, Lehmann (1959).

If the comparison of possible distributions for x is simple and direct, there will be a single transformation that carries any one parameter value into any other parameter value:

Assumption 3. *For any θ, θ' in Ω there is a single g in $\mathcal{G}$ such that $g^*\theta = \theta'$.* In terms of group-theory properties of $\mathcal{G}$ and $\mathcal{G}^*$, this can be expressed alternatively as follows: the mapping $g \to g^*$ is an *isomorphism* (that is, a 1-1 correspondence and this obtains if the identity transformation is the only transformation that does not change the distribution of the variable x); and the group $\mathcal{G}^*$ is *exactly transitive* on Ω (that is, for any θ, θ' there is a single g^* in Ω carrying θ into θ').

A sufficient statistic in Fisher's sense is an exhaustive statistic having the same dimensions as the parameter. Assumption 3 by simple analysis establishes isomorphisms between $\mathscr{G}, \mathscr{G}^*, \Omega$. The following assumption extends the isomorphisms to $\mathscr{X}$ and in this context gives a precise meaning to 'same dimensions':

Assumption 4. *$\mathscr{G}$ is exactly transitive on $\mathscr{X}$.*

These assumptions produce the structure described in § 2, and consequently the fiducial development in §§ 4 and 5. The assumptions are based on the use of a sufficient statistic and on my interpretation of *one aspect* of Fisher's essential ingredient. In this framework the fiducial development and interpretation in the preceding sections, and the fiducial properties in succeeding sections are remarkably coherent and consistent, and to me lend strong support to the pre-eminent position in which Fisher places fiducial inference.

7. Invariant measures

The specifications in the preceding sections have been described vaguely in terms of 'frequency distributions' or 'probability distributions'. For most problems of practical interest it is possible to describe frequency distributions by means of density functions with respect to a fixed or 'carrying' measure on the sample space. Here, with transformations on the space, the carrying measure would be transformed into other carrying measures on the space. Also, with inverse transformations belonging to the group, it follows that these different carrying measures would be *absolutely continuous* one with respect of another (one expressable by means of a density function with respect to another; see, for example, Halmos (1950)). For simplicity of representation this suggests the use of a measure whose value is invariant when a transformation in a group is applied to a set in the sample space. Measure theory (for example, Halmos (1950)) establishes the existence of such an *invariant measure* in a quite general setting.

The use of density functions also facilitates the argument in succeeding sections which are concerned with prior distributions and with the combination of fiducial distributions.

A measure μ on a group $\mathscr{G}$ is said (Halmos (1950), Loomis (1953)) to be a *left invariant* measure or a *left Haar* measure if it satisfies

$$\mu(G) \equiv \mu(gG)$$

for all elements g in $\mathscr{G}$ and all Borel sets G contained in $\mathscr{G}$; gG is the set obtained when each element in G is multiplied on the left by g. Such a measure exists if the group satisfies:

Assumption 5. *The group $\mathscr{G}$ has a locally compact metric topology and the group operations are continuous with respect to the metric.*

Measure theory establishes the essential uniqueness of left Haar measures, uniqueness in the sense that for any two left Haar measures one can be expressed as a constant times the other.

Let G^{-1} be the set of elements g^{-1} corresponding to elements g in G. Also, let ν be a measure on the group $\mathscr{G}$ defined by

$$\nu(G) = \mu(G^{-1}),$$

where μ is a left Haar measure. It is easily shown that ν has the property

$$\nu(G) = \nu(Gg)$$

for all g in $\mathscr{G}$ and all Borel sets G contained in $\mathscr{G}$. By analogy such a measure is called a *right invariant* or *right Haar* measure.

Let μ_g be defined by

$$\mu_g(G) = \mu(Gg);$$

μ_g is easily shown to be a measure. It satisfies the property of left invariance and hence by the uniqueness theorem must be a constant (which can of course depend on the g entering its definition) times the measure μ

$$\mu(Gg) = \Delta(g)\,\mu(G).$$

This positive real valued function $\Delta(g)$ defined on $\mathscr{G}$ is called the *modular function* of the group $\mathscr{G}$. Simple analysis establishes the following properties of the modular function

$$\Delta(gh) = \Delta(g)\,\Delta(h), \qquad \Delta(g^{-1}) = 1/\Delta(g),$$
$$\nu(gG) = \Delta(g^{-1})\,\nu(G), \quad \Delta(e) = 1,$$

where e is the identity element in the group.

The prime importance of the modular function is that it gives the relationship between left and right Haar measures—either can be expressed by means of a density function with respect to the other

$$\mu(G) = \int_G d\mu(g) = \int_G \Delta(g)\,d\nu(g),$$
$$\nu(G) = \int_G d\nu(g) = \int_G \Delta(g^{-1})\,d\mu(g),$$

or more briefly $d\mu = \Delta\,d\nu$, $d\nu = \Delta^{-1}\,d\mu$.

The example in § 2 was discussed in terms of location and scale transformations. These transformations form a group with elements $[a, b]$, where $-\infty < a < \infty$, $0 < b < \infty$. The left Haar measure for this group has element

$$\frac{da\,db}{b^2};$$

the invariance of this is easily checked. (The Jacobian of the *linear* transformation

$$[A, B] = [c, d]\,[a, b]$$

is d^2; also $B^2 = d^2b^2$.) The right Haar measure element is

$$\frac{da\,db}{b}$$

and the modular function is $\Delta([a, b]) = 1/b$.

8. The specification in terms of density functions

In this section a density function is assumed for the pivotal variable g and is used to produce expressions for the density functions of x given θ and of θ given x.

In § 2 it was shown that the sample space $\mathscr{X}$ could be identified with the group $\mathscr{G}$ by taking a reference point x_0 in $\mathscr{X}$. The identification then has the form

$$x \leftrightarrow g_x,$$

where g_x is the group element satisfying $x = g_x x_0$. Thus, in effect the sample space has the structure of the group $\mathscr{G}$. A simplification of notation is then obtained if x is used to designate the group element g_x. In a similar manner the parameter values can be replaced by group elements $\theta_0 \to e$, $\theta \to h_\theta$ and the notation then simplified by using θ to designate the group

element h_θ. Thus the sufficient statistic x, the transformation g, the parameter θ range over the spaces $\mathscr{X}$, $\mathscr{G}$, Ω which all have the structure of the group $\mathscr{G}$.

In this simplified notation the pivotal equation is

$$g = \theta^{-1}x$$

and its solutions are $x = \theta g$; $\theta = xg^{-1}$.

Suppose now that the sufficient statistic x has a distribution expressible by means of a density function with respect to left Haar measure. The distribution of the pivotal variable g is in fact the distribution of x for the case in which the parameter $\theta = e$; let $p(g)$ be the probability density function in this case:

$$\Pr\{g \text{ in } G\} = \int_G p(g)\, d\mu(g).$$

The general distribution of x can then be obtained from the pivotal equation $x = \theta g$.

$$\begin{aligned}
\Pr\{x \text{ in } X|\theta\} &= \Pr\{\theta g \text{ in } X\} \\
&= \Pr\{g \text{ in } \theta^{-1}X\} \\
&= \int_{\theta^{-1}X} p(g)\, d\mu(g) \\
&= \int_{x \text{ in } X} p(\theta^{-1}x)\, d\mu(\theta^{-1}x) \\
&= \int_X p(\theta^{-1}x)\, d\mu(x),
\end{aligned}$$

the last step following from the left invariant properties of μ. The essentials of this derivation can be exhibited in formal manipulation of probability elements. The variable x is generated from the variable g by the pivotal equation; the probability element is

$$p(g)\, d\mu(g) = p(\theta^{-1}x)\, d\mu(\theta^{-1}x) = p(\theta^{-1}x)\, d\mu(x).$$

Thus the density function for x given θ is $p(\theta^{-1}x)$ with respect to left Haar measure.

The pivotal equation solved for θ is $\theta = xg^{-1}$. The fiducial distribution for θ is generated by this equation from the variable g; x is kept fixed

$$\begin{aligned}
\Pr\{\theta \text{ in } \Theta|x\} &= \Pr\{xg^{-1} \text{ in } \Theta\} \\
&= \Pr\{g \text{ in } \Theta^{-1}x\} \\
&= \int_{\Theta^{-1}x} p(g)\, d\mu(g) \\
&= \int_{\theta \text{ in } \Theta} p(\theta^{-1}x)\, d\mu(\theta^{-1}x) \\
&= \int_\Theta p(\theta^{-1}x)\, \Delta(x)\, d\mu(\theta^{-1}) \\
&= \int_\Theta p(\theta^{-1}x)\, \Delta(x)\, d\nu(\theta).
\end{aligned}$$

The essentials of this derivation can be exhibited in formal manipulation of probability elements

$$p(g)\, d\mu(g) = p(\theta^{-1}x)\, d\mu(\theta^{-1}x) = p(\theta^{-1}x)\, \Delta(x)\, d\nu(\theta).$$

Thus the density function for θ given x is essentially $p(\theta^{-1}x)$ with respect to *right* Haar measure; $\Delta(x)$ is a normalizing constant to make the total integral equal to 1.

In a sense it is natural to use right Haar measure as a basis for θ and left Haar measure as a basis for g and for x. For, let G be a set of values for g and define X and Θ by

$$G = \theta^{-1}X \quad \text{for} \quad \theta \text{ fixed}$$
$$= \Theta^{-1}x \quad \text{for} \quad x \text{ fixed}.$$

Then
$$\mu(G) = \mu(\theta^{-1}X) = \mu(X)$$
$$= \mu(\Theta^{-1}x) = \nu(x^{-1}\Theta) = \Delta(x)\,\nu(\Theta).$$

Thus left Haar measure for x and g relates naturally to right Haar measure for θ.

9. Combining fiducial and prior distributions

Consider a specification of the kind in § 2 and suppose that the specification admits description in terms of density functions as in the preceding section. Also suppose that prior information concerning the parameter can be summarized in terms of an *a priori* frequency distribution with probability measure $N(\Theta)$, and that this admits a density function $n(\theta)$ with respect to the 'natural' right Haar measure. The *a priori* probability element for θ is then given by

$$n(\theta)\,d\nu(\theta).$$

The fiducial probability element for possible values of θ given x is

$$p(\hat{\theta}^{-1}x)\,\Delta(x)\,d\nu(\hat{\theta}).$$

The joint probability element for θ and $\hat{\theta}$ on $\Omega \times \Omega$ is

$$p(\hat{\theta}^{-1}x)\,\Delta(x)\,n(\theta)\,d\nu(\theta)\,d\nu(\hat{\theta}).$$

But θ and $\hat{\theta}$ need to be identified, since the prior distribution produced the θ that yielded the x observed; taking the conditional distribution along $\theta = \hat{\theta}$ with respect to the base measure ν yields the following relative density for θ given x

$$p(\theta^{-1}x)\,n(\theta)$$

which when normalized yields the probability element

$$\frac{p(\theta^{-1}x)\,n(\theta)\,d\nu(\theta)}{\int_\Omega p(\phi^{-1}x)\,n(\phi)\,d\nu(\phi)}.$$

It is interesting to note that the above 'logical' combination of fiducial and *a priori* distributions is just the ordinary *a posteriori* distribution for θ. The joint probability element for x and θ is

$$p(\theta^{-1}x)\,n(\theta)\,d\mu(x)\,d\nu(\theta)$$

which yields the following *a posteriori* probability element for θ given x

$$\frac{p(\theta^{-1}x)\,n(\theta)\,d\nu(\theta)}{\int_\Omega p(\phi^{-1}x)\,n(\phi)\,d\nu(\phi)}$$

which is the expression just derived by 'logical' combination.

Fisher's 'essential ingredient' in its entirety requires the absence of *all* prior information. In the present framework based on one aspect of Fisher's 'ingredient' there is no need to require the absence of an *a priori* distribution, since it can be combined with the fiducial information in a logical manner and be completely consistent with a Bayesian analysis.

10. Conditioning fiducial distributions

Suppose that the information available before an observation is taken on the sufficient statistic x is such as to restrict the range of possible values of the parameter to some subset Ω_0 of Ω. Fisher's essential ingredient, interpreted strictly, would then say that the fiducial argument was not applicable. Fisher himself, however, in an example in *Statistical Methods and Scientific Inference* (Fisher, 1956), has derived a fiducial distribution in such a situation. Within the framework introduced in this paper, a restriction on the range of the parameter seems entirely consistent with the use of the fiducial method.

The fiducial distribution for θ given x can be described by $\hat{\theta} = g_x g^{-1}\theta_0$, where g has the pivotal distribution. In the notation of Section 8 this produces the fiducial probability element

$$p(\hat{\theta}^{-1}x)\,d\nu(\hat{\theta}),$$

where ν is right Haar measure on $\mathscr{G}$. The argument in § 5 gave a frequency interpretation to this fiducial probability; it seems natural then to impose the condition 'θ in Ω_0' as a condition on the fiducial distribution to obtain a *conditioned fiducial distribution* with element

$$\frac{\phi_{\Omega_0}(\hat{\theta})\,p(\hat{\theta}^{-1}x)\,d\nu(\hat{\theta})}{\int \phi_{\Omega_0}(\eta)\,p(\eta^{-1}x)\,d\nu(\eta)},$$

where $\phi_{\Omega_0}(\hat{\theta}) = 1, 0$ according as $\hat{\theta}$ belongs to, does not belong to, Ω_0. This has a frequency interpretation by means of a straightforward restriction to Ω_0 of the frequency interpretation in the unrestricted case.

In the example on page 133 of *Statistical Methods and Scientific Inference*, Fisher has proceeded differently and condensed the fiducial probability outside the permissible region to a boundary point. Fisher gives no justification for his procedure.

11. Combining fiducial distributions

Consider two observational systems concerned with a single parameter θ, and suppose that the specifications have the form described in § 8. For one of the systems let x be the sufficient statistic, let g be the pivotal variable with density function $p(g)$, and let $\hat{\theta}_1$ be the fiducial variable derived from an observation on x. Similarly, for the second system let y be the sufficient statistic, let h be the pivotal variable with density function $q(h)$, and let $\hat{\theta}_2$ be the fiducial variable derived from an observation on y. In this section several methods are considered for combining the observational information; reassuringly, they yield identical results.

As a first method, consider the direct combination of the fiducial distributions. The joint probability element for $\hat{\theta}_1$ and $\hat{\theta}_2$ is

$$p(\hat{\theta}_1^{-1}x)\,q(\hat{\theta}_2^{-1}y)\,\Delta(x)\,\Delta(y)\,d\nu(\hat{\theta}_1)\,d\nu(\hat{\theta}_2).$$

The available prior information states that the parameter values in the two systems are equal. Formalizing this in the condition $\hat{\theta}_1 = \hat{\theta}_2$ and imposing this condition by reference to the natural right Haar measure on Ω produces the following 'fiducial' probability element for θ based on the combined systems

$$\frac{p(\hat{\theta}^{-1}x)\,q(\hat{\theta}^{-1}y)\,d\nu(\hat{\theta})}{\int_\Omega p(\eta^{-1}x)\,q(\eta^{-1}y)\,d\nu(\eta)}.$$

As a second method, consider the use of the fiducial distribution from one system as a prior distribution for a Bayesian argument on the other system. The fiducial density function from the first system is

$$p(\hat{\theta}^{-1}x)\Delta(x)$$

with respect to right Haar measure. Using this as a prior density $n(\theta)$ for a Bayesian argument on the second system (see § 9) produces an *a posteriori* probability element for θ

$$\frac{p(\hat{\theta}^{-1}x)\,q(\hat{\theta}^{-1}y)\,dv(\hat{\theta})}{\int_{\Omega} p(\eta^{-1}x)\,q(\eta^{-1}y)\,dv(\eta)}.$$

This is the result obtained by the first method.

For a third method, suppose that the combination of the two observational systems admits a sufficient statistic $r(x,y)$. The joint density function for x, y can then be factored

$$p(\theta^{-1}x)\,q(\theta^{-1}y) = f(r(x,y),\theta)\,w(x,y).$$

Let k be some value of the statistic $r(x,y)$ and let $t(x,y)$ be the solution of

$$r(t^{-1}x, t^{-1}y) = k;$$

moderate regularity conditions will insure a unique solution. The definition of the function $t(x,y)$ shows that $t(\delta^{-1}x, \delta^{-1}y)$ is a solution of

$$r(t^{-1}\delta^{-1}x, t^{-1}\delta^{-1}y) = k;$$

direct substitution shows that $\delta t(x,y)$ is also a solution of the equation. Hence

$$t(\delta^{-1}x, \delta^{-1}y) = \delta^{-1}t(x,y)$$

and $t(x,y)$ is transformed in the same way as a group element. The joint density function can be manipulated as follows

$$\begin{aligned} p(\theta^{-1}x)\,q(\theta^{-1}y) &= p((t^{-1}\theta)^{-1}\,t^{-1}x)\,q((t^{-1}\theta)^{-1}\,t^{-1}y) \\ &= f(r(t^{-1}x, t^{-1}y), t^{-1}\theta)\,g(t^{-1}x, t^{-1}y). \end{aligned}$$

Substituting $t = t(x,y)$ produces

$$\begin{aligned} p(\theta^{-1}x)\,q(\theta^{-1}y) &= f(k, t^{-1}(x,y)\,\theta)\,w(t^{-1}(x,y)\,x, t^{-1}(x,y)\,y) \\ &= f^*(\theta^{-1}t(x,y))\,w^*(x,y), \end{aligned}$$

where f^* is implicitly defined by the last step. This proves that $t(x,y)$ is a sufficient statistic. In addition, $t(x,y)$ takes values in the group $\mathcal{G}$ and transforms simply: $t(gx, gy) = gt(x,y)$. The last equation also establishes that $w^*(x,y)$ is invariant: $w^*(gx, gy) = w^*(x,y)$.

Simple analysis involving the last displayed equation, the invariance of $w^*(x,y)$, and the transformation properties of $t(x,y)$ show that the density function for t is $Kf^*(\hat{\theta}^{-1}t)$ with respect to left Haar measure (K is a normalizing constant). The corresponding fiducial probability element for θ given t is then

$$kf^*(\hat{\theta}^{-1}t)\,\Delta(t)\,dv(\hat{\theta}).$$

The fiducial probability element obtained by the first two methods can be simplified in the face of the assumptions and notation used for this third method

$$\frac{p(\theta^{-1}x)\,q(\theta^{-1}y)\,d\nu(\theta)}{\int_\Omega p(\eta^{-1}x)\,q(\eta^{-1}y)\,d\nu(\eta)} = \frac{f^*(\theta^{-1}t(x,y))\,w^*(x,y)\,d\nu(\theta)}{\int_\Omega f^*(\eta^{-1}t(x,y))\,w^*(x,y)\,d\nu(\eta)}$$

$$= \frac{f^*(\theta^{-1}t(x,y))\,\Delta(t(x,y))\,d\nu(\theta)}{\int_\Omega f^*(\eta^{-1}t(x,y))\,\Delta(t(x,y))\,d\nu(\eta)}$$

$$= kf^*(\theta^{-1}t(x,y))\,\Delta(t(x,y))\,d\nu(\theta).$$

Thus the first two methods yield the same result as the third in problems in which the third method is applicable.

In the third method the two observational systems are put together to form a single composite system and analysed under the assumption that the composite system admits a sufficient statistic. For a fourth method of analysis, suppose that the composite system does not admit a sufficient statistic. The probability element is

$$p(\theta^{-1}x)\,q(\theta^{-1}y)\,d\mu(x)\,d\mu(y)$$

on the product space $\mathscr{X} \times \mathscr{Y}$. This model has the more general form discussed in §3; an ancillary statistic is

$$a = x^{-1}y$$

and a conditionally sufficient statistic is

$$t = x.$$

The statistic $a = x^{-1}y = (fx)^{-1}fy$ is invariant under transformations in the group and hence has a fixed distribution—is ancillary. The statistic t on the other hand transforms as a group element under transformations in the group.

The carrying measure of a set $A \times T$ of values (a, t) is

$$M(A \times T) = \int_{\substack{x^{-1}y \text{ in } A \\ x \text{ in } T}} d\mu(x)\,d\mu(y)$$

$$M(A \times gT) = \int_{\substack{x^{-1}y \text{ in } A \\ x \text{ in } gT}} d\mu(x)\,d\mu(y)$$

$$= \int_{\substack{x^{-1}y \text{ in } A \\ x \text{ in } T}} d\mu(gx)\,d\mu(gy)$$

$$= \int_{\substack{x^{-1}y \text{ in } A \\ x \text{ in } T}} d\mu(x)\,d\mu(y)$$

$$= M(A \times T).$$

Thus the carrying measure as it applies to sets of values for t, given values for a, is left invariant and hence is a constant times the left Haar measure μ

$$M(A \times T) = K\mu(T)$$

for A fixed.

Making the change of variable in the original density function and using the conditional form of the carrying measure gives the following expression for the conditional probability element of t given a

$$Kp(\theta^{-1}t)\,q(\theta^{-1}ta)\,d\mu(t).$$

The fiducial distribution of θ given t (and a) is then

$$Kp(\theta^{-1}t)\,q(\theta^{-1}ta)\,\Delta(t)\,d\nu(\theta)$$

and this is the result obtained by the first two methods.

Four methods have been considered for obtaining fiducial distributions from separate systems concerned with the same parameter. The resulting distributions are identical, a very reassuring internal-consistency property of fiducial analysis in a transformation framework. Consider separate observational systems concerned with a single parameter, and suppose that each has the transformation form of §§ 2 or 3. The first two methods, consistent one with the other, give a 'reasonable' way of obtaining an overall fiducial distribution and the result is the same as obtained by the fourth method. Thus in effect the ancillary statistic for the overall system is *generated* by the 'reasonable' method of combining the individual fiducial distributions. In such situations then the ancillary statistic is not as arbitrary a thing as might be first thought, but is an intrinsic part of the model for purposes of inference.

If x and θ are real valued and if the assumptions in § 2 are fulfilled then the problem is of location parameter form. If the distributions are discrete this result follows simply. If the distributions are continuous the result follows from Theorem 2.9.3 in Cohn (1957).

Under mild regularity conditions Lindley (1958) has examined the consistency of methods 2 and 3 for combining fiducial distributions. In the case that the variables correspond to different observations from the same distribution having a real variable and real parameter, he has shown that the consistency is obtained if and only if the parameter is essentially a location parameter—that is, that the group structure herein considered is present for the real variables.

Consider again the location and scale parameter example mentioned in § 3: $x_1, \ldots, x_n$ is a sample from a distribution with density

$$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

on the real line; f is specified. The density function for a sample of n is

$$\frac{1}{\sigma^n} \prod_{i=1}^{n} f\left(\frac{x_i-\mu}{\sigma}\right).$$

For a fiducial analysis based on § 3 the conditional distribution of $(\bar{x}, s)$ given the ancillary statistic A is needed and this is obtained directly from the joint density of $\bar{x}, s, A$. The Jacobian of the transformation from $(x_1, \ldots, x_n)$ to $(\bar{x}, s, A)$ depends on $(\bar{x}, s)$ only in the form of a factor s^{n-2}. Thus the conditional probability element for $\bar{x}, s$ given A, μ, σ has the form

$$\frac{K}{\sigma^n} \prod_{i=1}^{n} f\left(\frac{x_i-\mu}{\sigma}\right) s^{n-2}\, s^2 \frac{d\bar{x}\,ds}{s^2},$$

where the measure element has been adjusted to exhibit left Haar measure. The fiducial method of § 5 together with the measure results of §§ 7 and 8 then give the following fiducial probability element for μ, σ given $\bar{x}, s, A$:

$$\frac{K}{\sigma^n} \prod_{i=1}^{n} f\left(\frac{x_i-\mu}{\sigma}\right) s^{n-2}\, s^2 \frac{1}{s} \frac{d\mu\,d\sigma}{\sigma}.$$

If $n = 2$ no ancillary statistic is needed and the pair of variables is a sufficient statistic. In this simple case the fiducial distribution for μ, σ can be derived without the seeming arbitrariness of an ancillary statistic.

For n greater than 2 (and for present simplicity suppose it is even) the set of variables can be broken into pairs and a fiducial distribution from each obtained. The general result of this section then demonstrates that an overall fiducial distribution obtained by direct combination (§ 11, method one) or by a Bayesian argument (§ 11, method two) will be just the fiducial distribution obtained here by the ancillary argument. This illustrates the intrinsic nature of the ancillary statistic for problems in the transformation framework, and gives an alternative demonstration of the essential uniqueness of the ancillary statistic for this example.

The discussion in this section has been restricted to specifications that satisfy the transformation conditions in §§ 2 and 3. The fiducial method has been applied to other kinds of specifications and Quenouille (1958) in his book, *The Fundamentals of Statistical Reasoning*, has discussed methods for combining fiducial distributions.

12. The Behrens–Fisher problem

Let $x_1, \ldots, x_m$ be a sample from a normal distribution (μ_1, σ_1^2), and let $y_1, \ldots, y_n$ be a sample from a normal distribution (μ_2, σ_2^2). The Behrens–Fisher problem is concerned with the estimation of the parameter difference $\mu_1 - \mu_2$.

The information concerning the frequency distributions relates the x's to (μ_1, σ_1^2) and the y's to (μ_2, σ_2^2). The transformation analysis in § 5 shows that this information is such that, given the observed values $\bar{x}, s_x^2$, the knowledge concerning μ_1 is that of a variable described by

$$\bar{x} + t_1 s_x / \sqrt{m},$$

where t_1 has Student's distribution with $m - 1$ degrees of freedom. Similarly, the information concerning μ_2 is that of a variable described by

$$\bar{y} + t_2 s_y / \sqrt{n},$$

where t_2 is a statistically independent variable with Student's distribution on $n - 1$ degrees of freedom. These distributions are appropriate to the observed values $\bar{x}, s_x, \bar{y}, s_y$. The frequency distribution of possible values for $\mu_1 - \mu_2$ is then described by

$$\bar{x} - \bar{y} + \left(t_1 \frac{s_x}{\sqrt{m}} - t_2 \frac{s_y}{\sqrt{n}} \right).$$

The distribution of this fiducial variable was derived by Fisher and percentage points for it can be obtained from tables prepared by Sukhatme.

This fiducial distribution for $\mu_1 - \mu_2$ has a frequency interpretation by means of the argument developed in § 5, and is relative to the fixed observed values $\bar{x}, \bar{y}, s_x, s_y$. It need not be and is not in general true that for a fixed value of $\mu_1 - \mu_2$, a 95 % fiducial interval will cover the value of $\mu_1 - \mu_2$ with frequency 95 %. Rather, for fixed $\bar{x}, \bar{y}, s_x, s_y$ the prior frequency information relating parameter to variable as it is given freedom in a transformation framework gives a frequency distribution of possible parameter values yielding the just-derived marginal distribution for $\mu_1 - \mu_2$. Some discussion on such a broader use of probability in science as opposed to technology may be found in Fisher (1955, 1959).

A fiducial interval for $\mu_1 - \mu_2$ will not be an invariant interval with respect to the transformations for the x's and for the y's. In fact, under separate linear transformations for the

x's and for the y's, the parameter $\mu_1-\mu_2$ itself is not transformed but is 'pulled apart'. A non-trivial invariant interval does not exist. However, the invariance was used only to establish the frequency distributions and in consequence taking a marginal distribution for $\mu_1-\mu_2$ need not be related to the invariance.

The referee who examined the original version of this paper offered many suggestions on arrangements and on detail which were very valuable to me in preparing the revision. I appreciate his help.

Prof. Lindley, as an editorial adviser for *Biometrika*, examined the paper in detail, made many valuable suggestions for reorganizing the material and thereby helped me towards the present version which I feel is a major improvement over the original. I thank him with deep appreciation.

REFERENCES

BUEHLER, ROBERT, J. (1959). Some validity criteria for statistical inferences. *Ann. Math. Statist*, **30**, 845–63.

COHN, P. M. (1957). *Lie Groups. Cambridge Tracts in Mathematics*, no. 46. Cambridge University Press.

FISHER, R. A. (1930). Inverse probability. *Proc. Camb. Phil. Soc.* **26**, 528–35 (reprinted in Fisher (1950) as paper 22).

FISHER, R. A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc.* A, **144**, 285–307 (reprinted in Fisher (1950) as paper 24).

FISHER, R. A. (1948). Conclusions fiduciaires. *Ann. Inst. Henri Poincaré*, **10**, 191–213.

FISHER, R. A. (1950). *Contributions to Mathematical Statistics.* New York: John Wiley and Sons, Inc.

FISHER, R. A. (1955). Statistical methods and scientific induction. *J. R. Statist. Soc. B*, **17**, 69–78.

FISHER, R. A. (1956). *Statistical Methods and Scientific Inference.* Edinburgh and London: Oliver and Boyd.

FISHER, R. A. (1959). Mathematical probability in the natural sciences. *Technometrics*, **1**, 21–30.

HALMOS, P. R. (1950). *Measure Theory.* New York: D. Van Nostrand.

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* New York: John Wiley and Sons, Inc.

LINDLEY, D. V. (1958). Fiducial distributions and Bayes' theorem. *J. R. Statist. Soc.* B, **20**, 102–7.

LOOMIS, L. H. (1953). *Abstract Harmonic Analysis.* New York: D. Van Nostrand.

MAULDON, J. G. (1955). Pivotal quantities for Wishart's and related distributions, and a paradox in fiducial theory. *J. R. Statist. Soc.* B, **17**, 79–85.

QUENOUILLE, M. H. (1958). *The Fundamentals of Statistical Reasoning.* London: Charles Griffin and Co.,

TUKEY, JOHN, W. (1957). Some examples with fiducial relevance. *Ann. Math. Statist.* **28**, 687–95.

WALLACE, DAVID L. (1959). Conditional confidence level properties. *Ann. Math. Statist.* **30**, 864–76.

Biometrika (1961), **48**, 3 *and* 4, *p.* 281
Printed in Great Britain

Bayesian sequential analysis

BY G. B. WETHERILL
Birkbeck College

1. DISTRIBUTIONS CLOSED UNDER SAMPLING

Suppose we have batches of items presented for acceptance inspection, and each batch is to be sentenced independently by a sequential procedure. Let the result of inspection of an item be a random variable x with a density function $\phi(x|\theta)$, where $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$. In the simplest case where items are classified as effective or defective, $\phi(x|\theta)$ is a binomial distribution with probability of a defective equal to θ.

In acceptance inspection, the prior knowledge is represented by the process curve, which is a distribution of θ specifying the relative frequency with which batches are produced with quality θ.

In general let the prior knowledge be represented by a density function which we shall call a parameter distribution, denoted $\xi(\theta|\alpha)$, depending on a known constant α (where α may be a vector). For convenience throughout §§ 1 to 3 of the paper, equations will be formulated for continuous density functions $\xi(\theta|\alpha)$ and $\phi(x|\theta)$. A fully rigorous discussion of the theory will not be attempted.

After observations $x_1, x_2, \ldots, x_n$, have been taken the posterior distribution of θ is

$$\frac{\xi(\theta|\alpha).\phi(x_1|\theta).\phi(x_2|\theta)\ldots\phi(x_n|\theta)}{\int \xi(\theta|\alpha).\phi(x_1|\theta).\phi(x_2|\theta)\ldots\phi(x_n|\theta)\,d\theta}.$$

Suppose now that this posterior distribution is of the same functional form as $\xi(\theta|\alpha)$, say $\xi(\theta|\beta)$. Our total knowledge at any point is then represented as a distribution of this type. We shall call a distribution possessing this property a distribution closed under sampling, with respect to observations having a distribution $\phi(x|\theta)$.

*Formal definition (due to Barnard)**

A parametric family $\xi(\theta|\alpha)$ is said to be closed under sampling with respect to observations having a density function $\phi(x|\theta)$, if for all $x_1, x_2, \ldots, x_n$, for which $\phi(x_i|\theta) \neq 0$, there exists a β such that

$$\xi(\theta|\beta) = \frac{\xi(\theta|\alpha).\phi(x_1|\theta).\phi(x_2|\theta)\ldots\phi(x_n|\theta)}{\int \xi(\theta|\alpha).\phi(x_1|\theta).\phi(x_2|\theta)\ldots\phi(x_n|\theta)\,d\theta}, \tag{1}$$

where $\beta = \beta(\alpha, x_1, x_2, \ldots, x_n)$. To make this clear we shall consider one or two examples.

(*a*) Consider inspection of batches by attributes, and suppose that items are of two kinds, effective and defective. Suppose that the probability of a defective in each batch is p, and that p varies from batch to batch according to the Beta distribution $kp^s(1-p)^t$. This is our

* Dr M. Stone has independently studied the idea of distributions closed under sampling in an unpublished Ph.D. thesis.

parameter density function representing our prior knowledge. The posterior distribution after observing r defectives in n items is

$$\frac{kp^s q^t . p^r q^{n-r}}{k\int_0^1 p^s q^t . p^r q^{n-r} dp} = k' p^{r+s} q^{n-r+t}, \quad \text{where} \quad q = 1-p,$$

which is also a Beta distribution. The family $\xi(p|\beta)$ is therefore the class of all Beta distributions $kp^{\beta_1} q^{\beta_2}$, where $\beta = \{\beta_1, \beta_2\}$, and $\beta_1 = r+s$, $\beta_2 = n-r+t$.

Thus a Beta distribution is closed under sampling with respect to binomially distributed observations. The total knowledge at any point in the sampling can be represented in a two-dimensional space with co-ordinates (β_1, β_2).

(*b*) Suppose that our observations are distributed according to any probability distribution $\phi(x|\theta)$, where θ varies from batch to batch, but may only take certain given values θ_i $(i = 1, 2, \ldots, k)$. Our prior probability is then $(a_1, a_2, \ldots, a_k)$, where $\sum_1^k a_i = 1$. The posterior distribution is of the same form $(b_1, b_2, \ldots, b_k)$, where

$$b_i = \frac{a_i \prod_j \phi(x_j|\theta_i)}{\sum_{i=1}^k a_i \prod_j \phi(x_j|\theta_i)} \quad \text{and} \quad \sum_1^k b_i = 1.$$

The set of values θ does not change, and sampling merely alters the set of probabilities a_i. Therefore a prior distribution, which takes its values at a finite set of points θ, is closed under sampling with respect to observations distributed according to any probability function $\phi(x|\theta)$. The family $\xi(\theta|\alpha)$ is the class of all discrete distributions where the random variable θ may only take on the values θ_i $(i = 1, 2, \ldots, k)$.

(*c*) If our observations are distributed according to the Poisson distribution

$$\phi(x|\theta) = e^{-\theta} . \theta^x / x!$$

and if our prior knowledge is given by the Beta-type distribution

$$\xi(\theta|\alpha) = K\theta^s (1-\theta)^t,$$

where $0 < \theta < 1$, and the case $t = 0$ corresponds to the improper distribution $k\theta^s$ over all positive θ: then after taking observations $x_1, \ldots, x_n$, the posterior distribution is

$$\frac{K\theta^s (1-\theta)^t . e^{-n\theta} . \theta^m / \Pi_j x_j!}{K\int \theta^s (1-\theta)^t . e^{-n\theta} . \theta^m d\theta / \Pi_j x_j!}, \quad \text{where} \quad m = \Sigma_j x_j,$$

$$= \frac{\theta^{s+m} (1-\theta)^t e^{-n\theta}}{\int \theta^{s+m} (1-\theta)^t e^{-n\theta} d\theta}.$$

This is not of the same form as $\xi(\theta|\alpha)$, and hence here we have an example of a distribution not closed under sampling. It can be made closed by replacing the Beta distribution by the probability density

$$k\theta^{s+x} (1-\theta)^\beta e^{-\gamma\theta} / x!$$

and if we use this as our prior distribution, we see that it is closed under sampling with respect to observations having a Poisson distribution.

The great simplification introduced by considering distributions closed under sampling is that the problem of sequential decision is reduced from dealing with a random walk in infinitely many dimensions, to being a random walk in finitely many.

2. The ξ-space

If we have a parameter distribution $\xi(\theta|\alpha)$, where α may be a vector, and this distribution is closed under sampling, we can define some space ξ, in which each member of the family of distributions of the type $\xi(\theta|\alpha)$ is represented by a point. Our total knowledge at any stage in the sampling is a point in this space.

For examples of ξ-space we refer to examples (a) and (b) in § 1. The ξ-space for (a) is a plane with axes β_1 and β_2 and that for (b) is a k-dimensional simplex with vertices at the unit points.

Now, assuming that we are dealing with a prior distribution which is closed under sampling, consider some form of sequential inspection in which there are only two possible terminal decisions—to accept or reject the batch under inspection. Suppose also we can define the losses associated with making the terminal decisions when the true quality of the batch is specified by some particular value of θ. We define the risk of a terminal decision at any point in the ξ-space to the expected value of the loss with respect to θ. One terminal decision is preferred to the other if the risk associated with making this terminal decision is less than the risk associated with making the other terminal decision. If the quality of the batches remains unaltered during inspection, then the risk associated with making a terminal decision depends only upon the posterior distribution function, and hence on the position in the ξ-space. We can therefore associate with each point in the ξ-space the risks of taking either of the two kinds of terminal decision, and depending upon which of these is the smaller at each point, mark the ξ-space out into two regions, one in which acceptance is preferred and the other in which rejection is preferred. Clearly, apart from possible discontinuities, the boundary between these regions is the locus of all points such that the risks of the two kinds of terminal decision are equal. This boundary will be referred to as the neutral boundary.

This division of the ξ-space into terminal decision regions does not taking sampling into account. In fact, we shall have extremes in the ξ-space where the expected quality of the batch is so good or so bad that it should be accepted or rejected outright, without any sampling. In between there will be a region of doubt where it pays to take further observations. We thus have three regions, a continuation region and two terminal decision regions, each defined uniquely by position in the ξ-space.

Four boundaries are of interest: the boundaries between the terminal decision regions, and the continuation region; the boundary between the two terminal decision regions, and the boundary at which the continuation region and the two-terminal decision regions meet. Clearly, some of these boundaries may not exist, for example, in a likelihood ratio sequential procedure, the two terminal decision regions do not meet. We proceed to derive equations for each of these four boundaries.

3. Equations for the boundaries

We shall refer to the two terminal decisions as decisions 1 and 2; the boundary in the ξ-space between the decision 1 (or 2) region and the other terminal decision region or the

continuation region, will be referred to as the decision 1 (or 2) boundary. Let the losses of taking decisions 1 and 2 when θ specifies the true quality be denoted $W_1(\theta)$ and $W_2(\theta)$, respectively, where the unit of losses is the cost of a sample, which is assumed constant.

We now consider the boundary where the three decision regions meet; this is often a point and will then be referred to as the meeting point. Interest centres in the meeting point because if it exists and can be determined, the next section describes a method by which, in principle, the optimum terminal decision boundaries can be worked out backwards from the meeting point.

The neutral boundary is the locus of points such that the decision risks (where risk is defined as expected loss) of taking either terminal decision are equal, and this is given by α satisfying

$$\int \xi(\theta|\alpha)\,[W_2(\theta) - W_1(\theta)]\,d\theta = 0. \tag{2}$$

Clearly, if the decision 1 and 2 boundaries meet, they will meet on the neutral boundary.

Suppose now that the decision 1 and 2 boundaries meet the neutral boundary at the same point, then at this point the risk of making a terminal decision is equal to the cost of one more observation, plus the terminal decision risks associated with this further observation (one more observation from the meeting point can only lead to a terminal decision). If at the meeting point we make decision 2 (the resulting equation is the same whichever terminal decision is used here), then for a given θ, the risk of a terminal decision at the meeting point is simply $W_2(\theta)$. After one more observation is taken, the terminal decision risks are

$$q(\alpha, \theta)\,W_1(\theta) + [1 - q(\alpha, \theta)]\,W_2(\theta),$$

where $q(\alpha, \theta)$ is the probability that for an α on the meeting point boundary, one observation leads to points in terminal decision 1 region. Thus for the meeting point, α should satisfy

$$\int [q(\alpha, \theta)\,W_1(\theta) + (1 - q(\alpha, \theta))\,W_2(\theta) + 1]\,\xi(\theta|\alpha)\,d\theta = \int \xi(\theta|\alpha)\,W_2(\theta)\,d\theta$$

which reduces to

$$\int q(\alpha, \theta)\,[W_2(\theta) - W_1(\theta)]\,\xi(\theta|\alpha)\,d\theta = 1. \tag{3}$$

Two points remain to be clarified, first, that the two terminal decision boundaries do in fact meet the neutral boundary at the same point, and secondly, that the locus of α satisfying equations (2) and (3) is a complete definition of the meeting point boundary. Both of these points can be established in the following way. First, we derive equations for the decision 1 and 2 boundaries. Simultaneous solution of both of these equations defines the boundary where the two terminal decision boundaries meet, and it will be seen that the resulting equations can be reduced to equations (2) and (3) above.

Let points on the decision 1 and 2 boundaries be denoted by α' and α'', respectively. For sampling starting from position α in the continuation region of the ξ-space, with given boundaries, and the value of θ specified, let $S(\alpha, \theta)$ and $A(\alpha, \theta)$ be the expected further sample size, and the probability that sampling eventually terminates with decision 1, respectively. For points α' and α'' on the terminal decision boundaries, $S(\alpha', \theta)$, $A(\alpha', \theta)$, etc., are defined to be the values obtained if sampling is continued, but otherwise with the same terminal decision boundaries.

Consider any point α' on the decision 1 boundary, then the risk of continuing sampling is equal to the risk of taking terminal decision 1 immediately, see Barnard (1954).

To decide immediately has risk $\int \xi(\theta|\alpha')\, W_1(\theta)\, d\theta.$

To continue sampling has risk

$$\int [S(\alpha', \theta) + (1 - A(\alpha', \theta))\, W_2(\theta) + A(\alpha', \theta)\, W_1(\theta)]\, \xi(\theta|\alpha')\, d\theta.$$

The equation for the value(s) of α' on the decision 1 boundary is obtained by equating these expressions,

$$\int \xi(\theta|\alpha')\, W_1(\theta)\, d\theta = \int [S(\alpha', \theta) + (1 - A(\alpha', \theta)) W_2(\theta) + A(\alpha', \theta)\, W_1(\theta)] \xi(\theta|\alpha')\, d\theta \tag{4}$$

which reduces to

$$\int [S(\alpha', \theta) + (1 - A(\alpha', \theta))\, \{W_2(\theta) - W_1(\theta)\}]\, \xi(\theta|\alpha')\, d\theta = 0. \tag{5}$$

Similarly for points α'' on the decision 2 boundary we obtain the equation

$$\int \xi(\theta|\alpha'')\, W_2(\theta)\, d\theta = \int [S(\alpha'', \theta) + A(\alpha'', \theta)\, W_1(\theta) + (1 - A(\alpha'', \theta))\, W_2(\theta)]\, \xi(\theta|\alpha'')\, d\theta \tag{6}$$

which reduces to $$\int [S(\alpha'', \theta) + A(\alpha'', \theta)\, \{W_1(\theta) - W_2(\theta)\}]\, \xi(\theta|\alpha'')\, d\theta = 0. \tag{7}$$

It was pointed out above that the meeting-point boundary will be defined by $\alpha = \alpha' = \alpha''$ simultaneously satisfying equations (5) and (7). If such a solution exists, then provided the next step from points on this boundary do not land back in the continuation region (see §§ 4 and 8 for discussion of this and other assumptions), we have $S(\alpha, \theta) = 1$ and $A(\alpha, \theta) = q(\alpha, \theta)$. On substituting these values in equations (5) and (7), equation (7) becomes equation (3), and equation (5) can be reduced to equation (2). Thus if a solution to equations (2) and (3) exists, then this solution defines the locus of points where both the terminal decision boundaries meet the neutral boundary.

Equations (5) and (7) can be transformed into each other by using the equation for the neutral boundary. Therefore, no points in the ξ-space exist where the neutral boundary meets one only of the terminal decision boundaries.

Equations (5) and (7) are difficult to solve because of the complex form of $S(\alpha, \theta)$ and $A(\alpha, \theta)$. On the other hand, the function $q(\alpha, \theta)$ is usually very simple, so that the neutral boundary and the meeting-point boundary can be determined. The next section describes the numerical procedure by which the terminal decision boundaries can be worked out backwards from the meeting-point boundary.

Equations (5) and (7) can usually be solved and used in the following special case. Consider points α' and α'' on the terminal decision boundaries which are such that they lead only to points on the meeting-point boundary or to terminal decision points. For such points, $S(\alpha', \theta)$ and $S(\alpha'', \theta)$ are both unity, $A(\alpha', \theta)$ and $A(\alpha'', \theta)$ are of the form $q(\alpha, \theta)$, so that for points on the decision 2 boundary equation (3) holds, with a similar equation for the decision 1 boundary.

4. Determination of the boundaries from the meeting-point boundary

If we are considering a point β in the ξ-space, then the terminal decision risk at this point is

$$R_d(\beta) = \min \left\{ \int \xi(\theta|\beta)\, W_1(\theta)\, d\theta,\ \int \xi(\theta|\beta)\, W_2(\theta)\, d\theta \right\} \tag{8}$$

and the continuation risk is

$$R_c(\beta) = 1 + \int\int \xi(\theta|\beta)\,\phi(x|\theta)\,R(\beta')\,dx\,d\theta, \tag{9}$$

where β' denotes the posterior distribution for a prior distribution specified by β and an observation x, and where $R(\beta)$ is the risk at β, defined

$$R(\beta) = \min\{R_c(\beta), R_d(\beta)\} \tag{10}$$

so that if $R(\beta) = R_c(\beta)$, β is a continuation point, and if $R(\beta) = R_d(\beta)$, β is a terminal decision point.

In this section we make the following assumptions (in § 8 these assumptions are discussed again in some detail).

Assumption A. Given any particular origin in the ξ-space, if any point can be reached in n steps, then it cannot be reached in any other than n steps.

Assumption B. Given any particular origin in the ξ-space, of all points which can be reached in n steps, there is at most one which satisfies equation (2) for the neutral boundary. (This excludes such cases as testing two normal populations with unequal variances. This assumption is not necessary, and is introduced here for simplification.)

Assumption C. If a meeting-point boundary exists, then for a given origin in the ξ-space the maximum of the sample sizes associated with points on this boundary has some finite value, say N. (See § 8 for a further explanation of this assumption.)

With regard to assumption C, no points of the subset designated by N—other than the one on the meeting-point boundary—can be continuation points. This follows, since for a continuation point, it must be possible to proceed sampling and reach either terminal decision region, and this implies the existence of a point on the meeting-point boundary for some $N' > N$. Therefore, we can write down all the risks for the points of the subset designated by N. Further, points of the subset designated by $(N-1)$ all lead to points of the subset N, the risks of which are known. We can therefore apply equations (8), (9), and (10) to all the points of the subset $(N-1)$, calculate their risks and classify them as decision or continuation points. This process can now be repeated for the points of the subset $(N-2)$ and so on.

This procedure is quite simple in practice, and in principle can always be done numerically, by the method known as dynamic programming, see Bellman (1957). A large amount of computing will often be required, but the problem is ideally suited to an electronic computer.

5. Example 1. The mixed binomial distribution

Suppose that the distribution $\phi(x|\theta)$ is a binomial distribution with a probability θ of a defective item, and suppose that the parameter distribution for θ is a k-point binomial distribution with $k \geqslant 3$, so that $\text{prob}\,(\theta = \theta_i) = a_i$ $(i = 1, 2, \ldots, k)$, where $\sum_1^k a_i = 1$. The terminal decisions are to accept or reject the batch under inspection, and these are denoted decisions 1 and 2 respectively. The loss associated with taking decision i when θ_j is true is written W_{ij}.

Equation (2) gives the neutral boundary

$$\sum_1^k a_i \theta_i^r (1-\theta_i)^{n-r} (W_{1i} - W_{2i}) = 0, \tag{11}$$

where n is the number of items inspected, and r the number of defectives found. Similarly, in equation (3) the function $q(\alpha, \theta_i)$ is the probability, given θ_i, that one more observation from the meeting point leads to acceptance. The next observation can be either a good item or a bad item, and both lead to terminal decision points. One more good item can only increase the risk of rejection and decrease the risk of acceptance, so that $q(\alpha, \theta_i) = 1-\theta_i$, and equation (3) is

$$\sum_1^k a_i \theta_i^r (1-\theta_i)^{n-r} [1+(1-\theta_i)(W_{1i}-W_{2i})] = 0. \tag{12}$$

Wetherill (1960) has pointed out that equations such as these take on a particularly simple form if we assume the prior distribution to be such that $\theta_i/(1-\theta_i) = \lambda^i$, for $i = 1, 2, \ldots, k$, and it is also shown that these equations are easily solved by Horner's method. If we make this assumption, equations (11) and (12) for the meeting point become

$$\sum_1^k a_i (1-\theta_i)^n x^i (W_{1i}-W_{2i}) = 0 \tag{13}$$

and

$$\sum_1^k a_i (1-\theta_i)^n x^i [1+(1-\theta_i)(W_{1i}-W_{2i})] = 0, \tag{14}$$

where $x = \lambda^r$.

Numerical example. Suppose that we have $\lambda = \frac{1}{3}$, so that $\theta_1 = \frac{1}{4}$, $\theta_2 = \frac{1}{10}$, $\theta_3 = \frac{1}{28}$; also suppose that $W_{1j} = \max\{0, 500(\theta_j - 0{\cdot}07)\}$ and $W_{2j} = \max\{0, 500(0{\cdot}07 - \theta_j)\}$, so that $W_{11} = 90$, $W_{12} = 15$, $W_{23} = 17{\cdot}14$; and take the values of a_i to be $a_1 = 0{\cdot}05$, $a_2 = 0{\cdot}25$, and $a_3 = 0{\cdot}70$.

For this example, equations (13) and (14) are quadratic equations, and they have a common solution for an x corresponding to $r \simeq 1{\cdot}75$, $n = 15$. (For the method of solution see Wetherill (1960).) Because of the discreteness of the problem, the last point in the continuation region must correspond to integral n and r both less than or equal to the values given for the meeting point, so that for this example, $r = 0$ or 1, and $n \leqslant 15$. We can therefore proceed by determining the position of the acceptance and rejection boundaries for $r = 1$. This can be done by the method outlined in § 4, or alternatively it is quite simple for the stage immediately prior to the meeting point, to employ equations (5) and (7).

Consider equation (5) for the acceptance boundary. When $r = 1$ one further observation leads to a terminal decision, so that $S(\alpha, \theta_i) = 1$, and $A(\alpha, \theta_i) = 1-\theta_i$, and the equation becomes

$$\sum_1^k a_i \theta_i^1 (1-\theta_i)^{n-1} [1+\theta_i(W_{1i}-W_{2i})] = 0,$$

from which the value of n giving the position of the acceptance boundary at $r = 1$ can be determined. At the rejection boundary for $r = 1$, one further good item will not always lead to acceptance (if there is a point in the continuation region), and therefore $S(\alpha, \theta_i) \geqslant 1$, and $A(\alpha, \theta_i) \leqslant 1-\theta_i$. When equality holds in these expressions, equation (7) for the rejection boundary reduces to equation (12), and it follows that for $r = 1$, all points for which the expression of equation (12) is positive are certainly rejection points. Any points for $r = 1$ not thus classified by these two boundaries, may be either continuation points or rejection points, and these can be checked numerically by the method given below, or by using an equation similar to (12) in which we may have to introduce a more complicated function for $S(\alpha, \theta_i)$. For this example ($n = 9, r = 1$) is the last continuation point, and ($n = 10, r = 1$) and ($n = 7, r = 1$), are acceptance and rejection points, respectively.

For the problem outlined above, the procedure for working back to find the optimum decision boundaries is as follows. Denote the posterior probability of θ_i at any point (n, r) by a'_i, so that at (n, r)

$$a'_i = a_i \theta_i^r (1-\theta_i)^{n-r} \Big/ \sum_1^k a_i \theta_i^r (1-\theta_i)^{n-r}. \tag{15}$$

(Actually we should write this as $a'_i(n, r)$, but it simplifies the notation to use simply a'_i.) The terminal decision risk at (n, r), $R_d(n, r)$, is defined

$$R_d(n, r) = \min \left\{ \sum_1^k a'_i W_{1i}, \sum_1^k a'_i W_{2i} \right\}. \tag{16}$$

The probability of a bad item at (n, r) is $\sum_1^k a'_i p_i = b(n, r)$, say. Then the risk at any point is

$$R(n, r) = \min \{R_d(n, r), R_c(n, r)\}, \tag{17}$$

where $R_c(n, r)$ is the continuation risk defined by

$$R_c(n, r) = 1 + b(n, r)\, R(n+1, r+1) + [1 - b(n, r)]\, R(n+1, r). \tag{18}$$

On using these equations for the above numerical example we have the following sequential scheme:

Accept the batch when sampling reaches $n = 9$, $r = 0$.

Reject the batch when sampling reaches $r = 1$, and $n \leqslant 7$.

If sampling reaches $n = 8$, $r = 1$, inspect another item, and reject if it is bad.

If sampling reaches $n = 9$, $r = 1$, inspect another item, reject if it is bad, and accept if it is good.

The total risk of this scheme is 4·96, as against the risk of 7·2 of accepting without sampling.

6. Example 2. A beta prior distribution

Suppose that the distribution $\phi(x|\theta)$ is a binomial distribution with a probability θ of a defective item, and that the parameter distribution for θ is a Beta distribution $\theta^{s-1}(1-\theta)^{t-1}/\beta(s, t)$. Let the decision losses associated with wrongly accepting or rejecting a batch be $k\,|\theta - \theta_0|$.

Denote the number of bad items found during inspection by y, and the number of good items found by x, then equations (5) and (7), or alternatively (2) and (3), reduce to

$$(y+s)\,(1-\theta_0)/\theta_0 = x+t, \tag{19}$$

$$(y+s+x+t+1)\,(y+s+x+t)\,k^{-1} = (x+t)\,[\theta_0(x+t+1) - (1-\theta_0)\,(y+s)]. \tag{20}$$

From these equations it follows that the meeting point is at (X, Y), where

$$Y = k\theta_0^2(1-\theta_0) - \theta_0 - s, \tag{21}$$

$$X = (Y+s)\,(1-\theta_0)/\theta_0 - t. \tag{22}$$

Champernowne (1953) has written a paper on the most economic boundaries for the Beta distribution, but the values he gives for the meeting points are very different from the values given by the above equations. Actually Champernowne tabulates boundaries, not meeting points, and many of the differences are accounted for when we realize that the

farthest point reachable in sampling may be considerably nearer the origin than the meeting point of the optimum boundaries.

Suppose the meeting point is at (x, y), then it may well be that at $(x, y-1)$ it pays to accept, and at $(x-1, y)$ it pays to reject, while $(x-1, y-1)$ is a continuation point. Then clearly sampling will never reach (x, y), for if $(x-1, y-1)$ is reached, the next item sampled will always lead to a decision. In order to find the last reachable point in sampling we examine the points immediately off the neutral line, towards the origin from the meeting point of the boundaries, to see when they start being continuation points. If we are examining to see if (x, y) is the last reachable point in sampling, we compare the continuation and terminal decision risks for $(x, y-1)$ and $(x-1, y)$; if one (or both) of these is a continuation point, then (x, y) can be reached in sampling.

For any point, the terminal decision and continuation risks can be compared by comparing the left- and right-hand sides of either equation (4) or (6) (or alternatively using the expressions of equations (5) and (7)). Now suppose (x, y) is a continuation point, and both $(x+1, y)$ and $(x, y+1)$ are terminal decision points, then for $(x-1, y)$, given that sampling is continued, the probability of acceptance is $(1-\theta_0)^2$, so that the expected further sample size is $1 . \theta + 2(1-\theta) = 2-\theta$. Thus for α corresponding to $(x-1, y)$, we have

$$A(\alpha, \theta) = (1-\theta)^2 \quad \text{and} \quad S(\alpha, \theta) = 2-\theta.$$

We now substitute into equation (6), and if the left-hand side is greater than the right-hand side then $(x-1, y)$ is a continuation point, otherwise it is a terminal decision 2 point. The point $(x, y-1)$ can be examined in a similar way, and we proceed in this manner, examining points immediately off the neutral line until we find one or both of such points to be continuation points. We shall then have determined the last reachable point in sampling, as described above.

An example has been worked out by this method for $\theta_0 = \frac{1}{2}, k = 200$, so that the boundaries and decision risks are symmetrical. Equations (21) and (22) give the meeting point as $X = Y = 23{\cdot}5$. The last reachable point in sampling is $(6, 6)$, since by the above method $(5, 6)$ and $(6, 5)$ were found to be the farthest off-neutral line points which are continuation points. Champernowne has given boundaries for this case, and they give the same point, $(6, 6)$, as the last reachable point (his boundaries cease here). For this example, the boundaries determined by my method agree with those given by Champernowne with the exception of $(2, 0)$ which he gives as a continuation point, while he gives $(0, 2)$ as a decision point (which is correct). Clearly, the boundaries must be symmetrical. This is the only difference I have found between my boundaries and Champernowne's.

7. Example 3. A normal prior distribution

Suppose that the distribution $\phi(x|\theta)$ is normal with a mean m and unit variance, and that the parameter distribution for m is normal with mean μ and known variance σ^2. The two decisions to be made are to accept (decision 1), or to reject (decision 2), the batch under consideration, and we shall take the loss function to be $\alpha |m - m_0|$.

Clearly, the normal parameter distribution is closed under sampling with respect to normally distributed observations. If a sample of n observations has a mean $\bar{x}_n$, the posterior distribution of m is well known to be normal with a mean $(n\sigma^2\bar{x}_n + \mu)/(n\sigma^2 + 1)$ and variance $\sigma^2/(1+n\sigma^2)$, and this will be written $N(\mu_n, \sigma_n)$.

The neutral line is therefore

$$\frac{\alpha}{\sigma_n\sqrt{(2\pi)}}\int_{-\infty}^{\infty}(m-m_0)\exp-\frac{1}{2}\left\{\frac{m-\mu_n}{\sigma_n}\right\}^2 dm = 0,$$

which is
$$\mu_n = (n\sigma^2\bar{x}_n+\mu)/(n\sigma^2+1) = m_0, \tag{23}$$

or
$$\bar{x}_n = m_0+(m_0-\mu)/n\sigma^2. \tag{24}$$

Thus for all n, given a position (μ_n, n) on the neutral line, the value of x_{n+1} required to remain on the neutral line at $(n+1)$ is m_0.

In order to determine the meeting point of the boundaries we must evaluate equation (7), and for this example $q(\alpha,\theta) = \Phi(m_0|m)$, where $\Phi(x|m)$ is the normal probability integral with mean m and unit variance. Equation (7) becomes

$$\frac{\alpha}{\sigma_n\sqrt{(2\pi)}}\int_{-\infty}^{\infty}(m-m_0)\,\Phi(m_0|m)\exp-\frac{1}{2}\left\{\frac{m-\mu_n}{\sigma_n}\right\}^2 dm = 1. \tag{25}$$

The meeting point must be on the neutral line (23), and using this, equation (25) can be written

$$\frac{\alpha\sigma_N}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty}\Phi(m_0|m)\,d\exp-\frac{1}{2}\left\{\frac{m-m_0}{\sigma_N}\right\}^2 = 1,$$

where N denotes the sample size at the meeting point. This equation is clearly

$$\frac{\alpha\sigma_N^2}{\sqrt{(2\pi)}\sqrt{(1+\sigma_N^2)}} = 1. \tag{26}$$

On substituting for σ_N^2, we have a quadratic equation for N,

$$\alpha^2\sigma^4 = 2\pi[1+(N+1)\sigma^2][1+N\sigma^2]$$

and the only root of this which can be positive is

$$N = \sqrt{\left(\frac{\alpha^2}{2\pi}+\frac{1}{4}\right)}-\frac{1}{2}-\frac{1}{\sigma^2}. \tag{27}$$

The meeting point is therefore specified by $\mu_N = m_0$, where N is given by (27). The sample mean $\bar{x}_N$ corresponding to μ_N is obtained from (24) and (27),

$$\bar{x}_N = m_0+(m_0-\mu)/\{\sigma^2(\tfrac{1}{2}\alpha^2+\tfrac{1}{4})-\tfrac{1}{2}\sigma^2-1\}. \tag{28}$$

When N is large, we have approximately

$$N \simeq \frac{\alpha}{\sqrt{(2\pi)}}-\frac{1}{\sigma^2} \tag{29}$$

hence
$$\bar{x}_N \simeq m_0+\sqrt{(2\pi)}\,(m_0-\mu)/\alpha\sigma^2. \tag{30}$$

These functions appear to be of the right form. The sample size at the meeting point, N, is an increasing function of the losses α, and also of σ^2 which represents the precision of the prior distribution. The effect of the parameter μ of the prior distribution on $\bar{x}_N$ decreases with increasing N, and also decreases with increasing variance σ^2.

8. Conditions and assumptions

It is clearly of interest to know what are the necessary and sufficient conditions that (i) the procedure for working back to find the optimum boundaries can be performed, and (ii) that a meeting point exists. We shall reconsider assumptions A, B and C of § 4, and also

consider a fourth assumption, which we shall label D, that no sequence of sampling can result in repeating a posterior distribution which has been reached before.

First, it is clearly not necessary for there to be a meeting point for (i) to be possible, as it is sufficient if there is a 'last reachable point' in the continuation region, in the sense described in example 2. Thus if the decision boundaries tend asymptotically to the neutral line, (i) will be possible for discrete distributions $\phi(x|\theta)$. Further, assumption A is not necessary either, for (i) is possible if sampling can jump a stage from, say, a point in the set n to a point in the set $(n+2)$; however, it must not be possible for a point to lead to points in previous subsets, and assumption D is sufficient to ensure this, but may not be necessary.

Assumption B is really irrelevant, the essential point being that once a meeting point has been determined, we wish to be able to write down the risks for all points in the subset containing the meeting point. If what has been described as the meeting point is in fact a line or surface containing a (possible infinite) number of points, then for (i) to be possible all that is necessary is that a 'last reachable point' exists. If the neutral line has two or more branches, (i) is possible provided sampling is closed on each branch.

Assumption C is introduced to cover the case where the ξ-space has at least three dimensions, so that the meeting-point boundary is a line or surface, and it may be that for a given origin in the ξ-space, the number of steps taken to reach this boundary is finite for only part of the ξ-space, and otherwise infinite. For this situation, unless a last reachable point exists, (i) will be possible for at most only parts of the ξ-space. For an example of this consider testing among three binomial populations with probabilities θ_0, θ_1 and θ_2, where $\theta_0 > \theta_1 > \theta_2$, and three associated terminal decisions. There are parts of the ξ-space in which θ_0 or θ_2 have very small prior probabilities, so that the sequential procedure reduces to a test of two binomial populations, and clearly for the two separate cases of this, sampling may be closed for one and not for the other. For examples 1, 2 and 3 of this paper, the ξ-space is two-dimensional, and the question does not arise.

Now let us consider examples 1, 2 and 3. For example 3, the posterior distribution has a variance which is a decreasing function of n, and assumptions A, B and D hold. For example 2, similar remarks apply and for both examples 2 and 3, there exist explicit formulae for the meeting points.

The k-point binomial distribution of example 1 is more complicated. Although assumptions A, B and D all hold, there is not necessarily a meeting point, see Vagholkar & Wetherill (1960) and Wetherill (1957). The situation can be given briefly as follows. Suppose the loss functions are $k|\theta_i - \theta_0|$, then if we are sampling from a population with θ close to θ_0, there can be a stage in the sampling where all θ_i's except those either side of θ_0 are negligible. In such a situation, the optimum boundaries must approximate to those for a two-point binomial, which are open and parallel. To exclude this possibility it is necessary to put a condition on the loss functions, to ensure that it is not worth sampling for the two-point binomial scheme based on these two particular ordinates (i.e. those on either side of p_0) see Vagholkar & Wetherill (1960) and Vagholkar (1955).

It is interesting to notice that the two-point prior distribution (e.g. $k = 2$ in example 1), is such that for nearly every possible prior distribution there is a sequence of sample values (n, r) which will repeat it. This is easily proved, for the posterior distribution is a simple function of the likelihood ratio, λ, where

$$\log \lambda = r \log (\theta_1(1-\theta_2)/\theta_2(1-\theta_1)) + n \log ((1-\theta_1)/(1-\theta_2))$$

and there is a sequence of values of (n, r) which yield any given value of λ, provided the logarithms on the right-hand side have a rational ratio.

Now if a closed sequence of points is possible in the ξ-space, sampling can enter a closed loop, the sample size can be infinite, and the boundaries must be open. It would appear, therefore, that assumption C is necessary for a meeting point to exist.

I conjecture that, with 'reasonable' loss functions, and continuous probability densities $\xi(\theta|\alpha)$, assumption D is both necessary and sufficient for a meeting point to exist. (By 'reasonable' loss functions, I mean loss functions which are either monotonically increasing or decreasing functions of θ.)

The discussion in this section is incomplete, but it is hoped that sufficient has been said to give a lead to further work.

9. GENERALIZATIONS

We have assumed throughout that we have only two terminal decisions, but there is no difficulty in extending the theory to three or more decisions. With three decisions, for instance, the ξ-space can be divided into four regions, and there are in general two neutral lines.

The restriction to prior distributions closed under sampling could be abandoned, but this would yield much more complex results which are more difficult to apply, and I suspect that there is not a corresponding gain in generality.

Sections 1 to 3 are based on work in my Ph.D. thesis, and I am greatly indebted to Prof. G. A. Barnard for his generous help throughout. I am also indebted to Prof. D. R. Cox for some helpful discussions during preparation of the paper, and to Prof. D. V. Lindley for some useful comments on an earlier draft of this paper.

REFERENCES

BARNARD, G. A. (1954). Sampling inspection and statistical decisions. *J. R. Statist. Soc.* Ser. B, **16**, 151–74.

BELLMAN, R. (1957). *Dynamic Programming.* Princeton University Press, Princeton, New Jersey.

CHAMPERNOWNE, D. G. (1953). The economics of sequential sampling procedures for defectives. *Appl. Statist.* **2**, 118–30.

VAGHOLKAR, M. K. (1955). Application of statistical decision theory to sampling inspection schemes. Ph.D. thesis, University of London.

VAGHOLKAR, M. K. & WETHERILL, G. B. (1960). The most economical binomial sequential probability ratio test. *Biometrika*, **47**, 103–9.

WETHERILL, G. B. (1957). Application of the theory of decision functions to sampling inspection with special reference to cost of inspection. Ph.D. thesis, University of London.

WETHERILL, G. B. (1960). Some remarks on the Bayesian solution of the single sample inspection scheme. *Technometrics*, **2**, 341–52.

Biometrika (1961), **48**, 3 *and* 4, *p.* 293
Printed in Great Britain

Some properties of the Spearman estimator in bioassay*

BY BYRON WM BROWN, JR
School of Public Health, University of Minnesota, Minneapolis

1. INTRODUCTION AND SUMMARY

The quantal bioassay is one of the most frequent of biological experiments. There is need for simple analyses of the data from such experiments. The non-parametric Spearman procedures are satisfactory from the point of view of simplicity. Finney (1950, 1952*a*) has done some computations demonstrating that the bias of the Spearman estimator is small and the efficiency is high, if the dose-response function is normal or logistic and the doses are chosen close together and covering a wide range. Because the Spearman estimator can be computed without specifying the functional form of the dose-response function, it is desirable to know the properties of this estimator over a wide class of functions. In this paper Finney's experiment is formalized as an 'infinite experiment', the Spearman estimator is defined for this experiment, and bounds on the bias of the estimator are obtained over specified broad classes of dose-response functions. The variance of the estimator is approximated and this approximation is investigated. The high efficiency reported by Finney is verified for the normal and logistic functions, and high efficiencies are computed for a number of other functions, but a family of functions is found for which the efficiency can be arbitrarily close to zero.

2. SPEARMAN ESTIMATION AND PARAMETRIC ESTIMATION IN QUANTAL BIOASSAY

In the usual quantal bioassay model, the mean (or median) of a distribution function, $F(x)$, is estimated by performing the following experiment. k values are specified on the x-scale, $x_1, x_2, \ldots, x_k$. For each x_i, n_i observations are taken on the random variable with distribution $F(x)$. All that is observed is the number (r_i) of these n_i observations that do not exceed x_i. The $n_1+n_2+\ldots+n_k$ observations are assumed to be mutually independent. The result of the experiment is a set of mutually independent binomially distributed random variables, r_i, with means $n_i F(x_i)$ ($i = 1, 2, \ldots, k$).

A practical application of this model is in drug assay where $F(x)$ may be the probability of death for an animal given dose x (x may be a transform of the dose such as log dose). The purpose of the experiment may be to estimate the dose which will give 50 % probability of death. The experiment consists of giving varying doses of the drug to groups of animals and recording the number of animals that die in the several dose groups.

The usual estimation procedure (see Finney, 1952*b*) depends on a parametric representation of $F(x)$. The median or mean is estimated as a function of the maximum likelihood estimates of the unknown parameters. Berkson (1955) has recommended the use of other parametric estimators asymptotically equivalent to the maximum likelihood estimates but easier to compute.

* Part of a doctoral dissertation prepared under the joint guidance of E. A. Johnson and I. R. Savage, with financial support through a NIH Research Training Grant.

If the dose levels (x_i) are $x_1, x_2 = x_1+d, \ldots, x_k = x_1+(k-1)d$, the non-parametric Spearman estimator is defined as follows (Spearman, 1908):

$$\bar{x} = p_1(x_1-\tfrac{1}{2}d) + \sum_{i=1}^{k-1}(x_i+\tfrac{1}{2}d)(p_{i+1}-p_i)+(1-p_k)(x_k+\tfrac{1}{2}d)$$

$$= x_k+\tfrac{1}{2}d-d\sum_{i=1}^{k}p_i, \quad \text{where} \quad p_i = r_i/n_i. \tag{2·1}$$

The Spearman estimator compares favourably with the usual parametric estimators. It is simpler to compute, simpler in concept and non-parametric. Furthermore, exact distributional theory can be developed for it, and its efficiency is high. Results presented in this paper substantiate the last two points.

3. The infinite experiment and the Spearman estimator

The characteristics of the Spearman estimator are investigated here for an experiment involving an infinity of dose levels, $x_i = x_0+id$ $(i = 0, \pm 1, \pm 2, \ldots)$. The dose-mesh location, x_0, and the dose interval, d, will be assumed arbitrarily chosen and fixed for the present. At each dose level, x_i, n observations will be taken. All observations are assumed mutually independent.

The information for the binomial variable at dose level x_i is $n_iF_i(1-F_i)$. The information at all but a finite number of dose levels is negligible. Therefore, results obtained in investigating the infinite experiment will be relevant to finite experiments with dose levels covering the range of information.

For this infinite experiment the Spearman estimator will be taken from (2·1) to be

$$\bar{x} = \sum_{i=-\infty}^{\infty}(x_i+\tfrac{1}{2}d)(p_{i+1}-p_i), \tag{3·1}$$

or

$$\bar{x} = x_0+\tfrac{1}{2}d+d\sum_{i=1}^{\infty}(1-p_i)-d\sum_{i=-\infty}^{0}p_i. \tag{3·2}$$

4. The bias of the Spearman estimator for fixed infinite dose mesh

If $F(x)$ has a first moment μ, the series defining the estimator (3·1) converges with probability one and has mean

$$E(\bar{x}|x_0) = \sum_{-\infty}^{\infty}(x_i+\tfrac{1}{2}d)(F_{i+1}-F_i), \tag{4·1}$$

where $F_i = F(x_i)$. The bias of $\bar{x}$, conditional on x_0 and d, is exactly analogous to the bias due to grouping in the usual continuous distribution. The bias will be denoted by $B(\bar{x}|x_0)$ and can be written

$$B(\bar{x}|x_0) = E(\bar{x}|x_0)-\mu$$

$$= \sum_{i=-\infty}^{\infty}(x_i+\tfrac{1}{2}d)(F_{i+1}-F_i)-\int_{-\infty}^{\infty}x\,dF(x)$$

$$= \sum_{i=-\infty}^{\infty}(x_i+\tfrac{1}{2}d-c_i)(F_{i+1}-F_i), \tag{4·2}$$

where c_i is defined as follows

$$c_i = \int_{x_i}^{x_{i+1}}x\,dF(x)\Big/\int_{x_i}^{x_{i+1}}dF(x). \tag{4·3}$$

Since $x_i \leqslant c_i \leqslant x_{i+1}$, expression (4·2) implies the following lemma.

LEMMA 4·1. $|B(\bar{x})|x_0)| \leqslant \frac{1}{2}d$.

It is obvious that for a one-point distribution with mass point located at one of the dose levels, this dose level will be the mean, while the Spearman estimator will have value $\frac{1}{2}d$ units lower. Thus, the Spearman estimator attains the bound of $\frac{1}{2}d$. From this example, it is clear that for any $F(x)$, if d is large enough, the bias can approach $\frac{1}{2}d$.

In the following lemmas, tighter bounds on the bias are obtained and evaluated by imposing restrictions on the class of dose-response functions. In Lemma 4·2, $F(x)$ is assumed to have $s+1$ derivatives and symmetrical density function. In Lemma 4·3, $F(x)$ is assumed to have a unimodal density. Sharpness of the latter bound is shown.

LEMMA 4·2. Let $f(x)$ be the first derivative of $F(x)$. If $f(x)$ exists, is symmetrical, is differentiable s times, and if $f^{(v)}(x)$ has limit zero as $x \to \pm\infty$, for $v = 0, 1, 2, \ldots, s$, then

$$|B(\bar{x}|x_0)| \leqslant \sup_x |P_{v+1}(x)|\, d^{v+1} \int_{-\infty}^{\infty} |f^{(v)}(x)|\, dx \quad \text{for} \quad v = 0, 1, 2, \ldots, s, \tag{4·4}$$

where $P_v(x)$ is the vth Bernoulli function

$$P_{2u}(x) = \sum_{k=1}^{\infty} \frac{\cos 2k\pi x}{2^{2u-1}(k\pi)^{2u}} \quad (u = 1, 2, \ldots), \tag{4·5}$$

$$P_{2u+1}(x) = \sum_{k=1}^{\infty} \frac{\sin 2k\pi x}{2^{2u}(k\pi)^{2u+1}} \quad (u = 0, 1, 2, \ldots). \tag{4·6}$$

Proof. Let $\bar{x}_k$ be the Spearman estimator computed from the $2k+1$ dose levels about x_0:

$$\begin{aligned} \bar{x}_k &= \sum_{i=-k}^{k-1} (x_i + \tfrac{1}{2}d)(p_{i+1} - p_i) + p_{-k}(x_{-k} - \tfrac{1}{2}d) + (1 - p_k)(x_k + \tfrac{1}{2}d) \\ &= x_0 - d \sum_{-k}^{k} (p_i - \tfrac{1}{2}). \end{aligned} \tag{4·7}$$

Then the expected value of $\bar{x}_k$, conditional on x_0, d and n, and denoted by $E(\bar{x}_k|x_0)$, is as follows

$$E(\bar{x}_k|x_0) = x_0 - d \sum_{-k}^{k} [F(x_0 + id) - \tfrac{1}{2}]. \tag{4·8}$$

By the Euler–MacLaurin sum formula (see Cramer, 1951)

$$\begin{aligned} E(\bar{x}_k|x_0) = x_0 - d\Bigg\{ &\int_{-k}^{k} [F(x_0 + xd) - \tfrac{1}{2}]\, dx + \tfrac{1}{2}[F(x_0 + kd) - 1 + F(x_0 - kd)] \\ &- d \int_{-k}^{k} P_1(x) f(x_0 + xd)\, dx \Bigg\}. \end{aligned} \tag{4·9}$$

The expected value of $\bar{x}$, computed from the results over the infinite dose mesh, can be obtained by taking the limit of $E(\bar{x}_k|x_0)$ as k goes to infinity. This limit can be obtained by taking limits of the several terms in (4·9). The limit of the second term in brackets in (4·9) is zero, since F is a distribution function. The first term in brackets in (4·9) can be written

$$\int_{-k}^{k} [F(x_0 + xd) - \tfrac{1}{2}]\, dx = \frac{1}{d} \int_{\mu - kd + (x_0 - \mu)}^{\mu + kd + (x_0 - \mu)} [F(y) - \tfrac{1}{2}]\, dy. \tag{4·10}$$

Since $F(y)$ is symmetric, $F(y) - \frac{1}{2}$ is an odd function with respect to $y = \mu$, and the integral from $\mu - kd + (x_0 - \mu)$ to $\mu + kd - (x_0 - \mu)$ is zero. Therefore, it follows that

$$\int_{-k}^{k} [F(x_0 + xd) - \tfrac{1}{2}]\, dx = \frac{1}{d} \int_{\mu + kd - (x_0 - \mu)}^{\mu + kd + (x_0 - \mu)} [F(y) - \tfrac{1}{2}]\, dy. \tag{4·11}$$

Then, since F is a distribution function, it follows from (4·11) that

$$\lim_{k\to\infty}\int_{-k}^{k}[F(x_0+xd)-\tfrac{1}{2}]\,dx=\frac{x_0-\mu}{d}. \tag{4·12}$$

Using (4·12) and the limit of zero for the second term in brackets in (4·9), it follows that

$$\begin{aligned}E(\bar{x}|x_0)&=\lim_{k\to\infty}E(\bar{x}_k|x_0)\\&=\mu+\lim_{k\to\infty}d^2\int_{-k}^{k}P_1(x)f(x_0+xd)\,dx.\end{aligned} \tag{4·13}$$

Making use of the relationship between Bernoulli functions

$$P'_v(x)=(-1)^{v-1}P_{v-1}(x), \tag{4·14}$$

the integral in (4·13) can be integrated by parts to obtain

$$E(\bar{x}|x_0)=\mu+(-1)^v d^{v+1}\int_{-\infty}^{\infty}P_{v+1}\left(\frac{x-x_0}{d}\right)f^{(v)}(x)\,dx. \tag{4·15}$$

From (4·15) the bound on the bias, (4·4), can be obtained.

For $v=0$ and $v=1$, the bound (4·4) is

$$|B(\bar{x}|x_0)|\leqslant\tfrac{1}{2}d \qquad (v=0), \tag{4·16}$$

$$|B(\bar{x}|x_0)|\leqslant\tfrac{1}{8}d^2f_m \quad (v=1), \tag{4·17}$$

where

$$f_m=\max_x f(x).$$

If $f(x)$ has two points of inflexion, (4·4) for $v=2$ is simply written

$$|B(\bar{x}|x_0)|\leqslant 0{\cdot}032d^3f_1^{(1)}, \tag{4·18}$$

where $f_1^{(1)}$ is the absolute value of $f^{(1)}(x)$ at its points of inflexion.

The bounds on the bias, (4·4), depend on the assumption of symmetry of $f(x)$. The bound for $v=1$, (4·17), can be improved slightly if the assumption of unimodality of $f(x)$ is substituted for the assumption of symmetry.

LEMMA 4·3. If F has first derivative $f(x)$ unimodal with $f_m=\max_x f(x)$, then

$$|B(\bar{x}|x_0)|\leqslant\tfrac{1}{8}d^2f_m. \tag{4·19}$$

Proof. In (4·2) the ith term in the expression for the bias, $B(\bar{x}|x_0)$, will be denoted by B_i. This term can be written

$$\begin{aligned}B_i&=(x_i+\tfrac{1}{2}d-c_i)\,(F_{i+1}-F_i)\\&=\int_{x_i}^{x_{i+1}}(x_i+\tfrac{1}{2}d-x)f(x)\,dx.\end{aligned} \tag{4·20}$$

The unimodality of $f(x)$ implies that $f(x)$ is non-decreasing for $x\leqslant x_m$, and non-increasing for $x\leqslant x_m$, where x_m is the mode. Using this property it can be shown that the maximum absolute value of B_i, for $F_{i+1}-F_i$ fixed, is less than the value obtained for the step function, $g(x)$, with one jump appropriately placed in each interval, $x_i\leqslant x\leqslant x_{i+1}$, to yield the probability $F_{i+1}-F_i$ for the interval. For $x_{i+1}\leqslant x_m$

$$g(x)=f(x_i) \qquad (x_i\leqslant x\leqslant x_i+R_id),$$

$$g(x)=f(x_{i+1}) \quad (x_i+R_id\leqslant x\leqslant x_{i+1}),$$

where R_i is determined by

$$\int_{x_i}^{x_{i+1}} g(x)\,dx = F_{i+1} - F_i.$$

For $x_i \geqslant x_m$, the function $g(x)$ would be defined similarly. For the interval containing x_m, say (x_0, x_1), $g(x)$ would have a jump in (x_0, x_m) and another in (x_m, x_1) with the jumps located to yield appropriate probabilities, $F_m - F_0$ and $F_1 - F_m$, for the intervals.

Given the function $g(x)$, the values of the bias terms (4·20) for $g(x)$ can be obtained, and these will be bounds for the B_i computed from $f(x)$. For $x_{i+1} \leqslant x_m$ the result is

$$B_i \leqslant \tfrac{1}{2}d^2 R_i(1-R_i)(f_{i+1}-f_i). \tag{4·21}$$

Since $0 \leqslant R_i \leqslant 1$, it follows that $$B_i \leqslant \tfrac{1}{8}d^2(f_{i+1}-f_i). \tag{4·22}$$

Similar bounds for the intervals above the mode and for the interval containing the mode can be obtained. When these bounds are summed, the result is independent of the f_i and is the bound for the bias given in (4·19).

The bound (4·19) on the bias of the Spearman estimator is sharp. This is easily verified for $f(x)$ the uniform distribution and appropriate choice of x_0 and d.

Table 1. *Functional forms commonly used for $F(x)$*

Name	Functional form	Range	Variance
Logistic	$[1+e^{-\beta(x-\mu)}]^{-1}$	$-\infty < x < \infty$	$\dfrac{\pi^2}{3\beta^2}$
Normal	$\dfrac{1}{\sqrt{(2\pi)}}\displaystyle\int_{-\infty}^{\beta(x-\mu)} e^{-\frac{1}{2}t^2}\,dt$	$-\infty < x < \infty$	$\dfrac{1}{\beta^2}$
Angular	$\sin^2[\beta(x-\mu)+\frac{1}{4}\pi]$	$-\frac{1}{4}\pi < \beta(x-\mu) < \frac{1}{4}\pi$	$\dfrac{\pi^2-8}{16\beta^2}$
Uniform	$\beta(x-\mu)+\frac{1}{2}$	$-\frac{1}{2} < \beta(x-\mu) < \frac{1}{2}$	$\dfrac{1}{12\beta^2}$
One-particle or extreme value	$1-\exp(-e^{x-\gamma-\mu})$	$-\infty < x < \infty$	$\frac{1}{6}\pi^2$

Notes. (i) The means of all five distributions are μ. (ii) For the first four distributions, $\beta > 0$. (iii) For the fifth distribution, γ denotes Euler's constant.

Table 2. *Bounds for the bias of the Spearman estimator for the infinite experiment*

Tolerance distribution	Bound for the bias*	
	As a proportion of σ†	As a proportion of R‡
Logistic	$0{\cdot}0567\rho^2\sigma$†	$0{\cdot}0866\rho'^2R$‡
Normal	$0{\cdot}0499\rho^2\sigma$	$0{\cdot}0839\rho'^2R$
Angular	$0{\cdot}0427\rho^2\sigma$	$0{\cdot}0804\rho'^2R$
Uniform	$0{\cdot}0361\rho^2\sigma$	$0{\cdot}0750\rho'^2R$
One-particle or extreme value	$0{\cdot}1013\rho^2\sigma$	$0{\cdot}1561\rho'^2R$

* Using (4·19) for the bound, $|B(\bar{x}|x_0)| \leqslant (\frac{1}{8}d^2)f_m$.
† σ is the standard deviation of the tolerance distribution and $d = \rho\sigma$.
‡ R is the distance between the 20th and the 80th percentile and $d = \rho' R$.

Some frequently used functional forms for $F(x)$ are given in Table 1. In Table 2 the value of the bound (4·19) is given for each of the functions of Table 1. From Table 2 it can be seen that for these five functions the bias will be less than $0{\cdot}10\sigma$ for d less than σ. In keeping

with the non-parametric spirit of this investigation, and for purposes of quicker comprehension and approximation from actual data, the bounds in Table 2 (and Table 4) are presented also in terms of the difference (R) between the twentieth and eightieth percentiles. For the five distributions of Table 1, d less than $\frac{1}{2}R$ will result in a bias less than 4 % of R. In practice such a bias would be negligible.

5. THE VARIANCE OF THE SPEARMAN ESTIMATOR FOR FIXED INFINITE DOSE MESH

If $F(x)$ has a first moment μ, then, for fixed x_0, the series defining the Spearman estimator, (3·1), converges with probability one and the resulting estimate has variance $V(\bar{x}|x_0)$, given by

$$V(\bar{x}|x_0) = \frac{d^2}{n} \sum_{-\infty}^{\infty} F(x_i)\,[1-F(x_i)]. \tag{5·1}$$

Let $V(\bar{x}|x_0)$ be approximated by the integral

$$\overline{V}(\bar{x}) = \frac{d}{n} \int_{-\infty}^{\infty} F(1-F)\,dx. \tag{5·2}$$

LEMMA 5·1. For all F, x_0, d and n

$$|V(\bar{x}|x_0) - \overline{V}(\bar{x})| \leqslant \frac{d^2}{4n}. \tag{5·3}$$

Proof. Let x_m be a median of $F(x)$. Then $F(1-F)$ is non-decreasing for $x \leqslant x_m$ and non-increasing for $x \geqslant x_m$. Number the x_i so that $x_0 \leqslant x_m \leqslant x_1$. It follows that

$$\left.\begin{aligned} &\int_{x_{i-1}}^{x_i} F(1-F)\,dx \leqslant F_i(1-F_i)\,d \quad (i = 0, -1, -2, \ldots), \\ &\int_{x_i}^{x_{i+1}} F(1-F)\,dx \leqslant F_i(1-F_i)\,d \quad (i = 1, 2, 3, \ldots), \\ &\int_{x_0}^{x_1} F(1-F)\,dx \leqslant \tfrac{1}{2}\cdot\tfrac{1}{2}\,d. \end{aligned}\right\} \tag{5·4}$$

Combining the inequalities of (5·4) it follows that

$$d \sum_{-\infty}^{\infty} F_i(1-F_i) + \tfrac{1}{4}d \geqslant \int_{-\infty}^{\infty} F(1-F)\,dx. \tag{5·5}$$

Multiplying the inequality (5·5) by d/n yields

$$V(\bar{x}|x_0) - \overline{V}(\bar{x}) \geqslant -\frac{d^2}{4n}. \tag{5·6}$$

Inequalities analogous to those in (5·4) in the other direction can be written down and from them it can be seen that $\frac{1}{4}d^2/n$ is also an upper bound for the difference, $V(\bar{x}|x_0) - \overline{V}(\bar{x})$. The bound (5·3) is sharp. This is easily verified for $F(x)$ a two-point distribution with equal probabilities at the two points, and with d and x_0 appropriately chosen.

Further bounds in higher powers of d can be obtained using Euler–MacLaurin expressions for the variance (5·1). One result, for $F(x)$ symmetrical with two points of inflexion, is

$$|V(\bar{x}|x_0) - \overline{V}(\bar{x})| \leqslant \frac{d^3}{3n} f_m, \quad \text{where} \quad f_m = \max_x f(x). \tag{5·7}$$

Table 3 gives the values of $\overline{V}(\bar{x})$ for the distributions of Table 1. Table 4 gives the bounds on the relative deviation of $V(\bar{x}|x_0)$ from $\overline{V}(\bar{x})$, derived from (5·3) and (5·7), for the distributions of Table 1.

Table 3. *Values of $\overline{V}$ for several distributions*

Distribution	$\overline{V}(\bar{x})$*	
Logistic	$0{\cdot}5513\rho\frac{\sigma^2}{n}$†	$1{\cdot}2881\rho'\frac{R^2}{n}$‡
Normal	$0{\cdot}5642\rho\frac{\sigma^2}{n}$	$1{\cdot}5983\rho'\frac{R^2}{n}$
Angular	$0{\cdot}5750\rho\frac{\sigma^2}{n}$	$2{\cdot}0376\rho'\frac{R^2}{n}$
Uniform	$0{\cdot}5774\rho\frac{\sigma^2}{n}$	$2{\cdot}4942\rho'\frac{R^2}{n}$
One-particle or extreme value	$0{\cdot}5404\rho\frac{\sigma^2}{n}$	$0{\cdot}3508\rho'\frac{R^2}{n}$

* $\overline{V}(\bar{x}) = \frac{d}{n}\int_{-\infty}^{\infty} F(1-F)\,dx.$

† σ is the standard deviation of the tolerance distribution and $d = \rho\sigma$.

‡ R is the distance between the 20th and the 80th percentile and $d = \rho' R$.

Table 4. *Bounds for the relative deviation of* $|V(\bar{x}|x_0) - \overline{V}(\bar{x})|$

	Bound on $\frac{\lvert V(\bar{x}\lvert x_0) - \overline{V}(\bar{x})\rvert}{\overline{V}(\bar{x})}$					
Distribution	First bound*			Second bound*		
Logistic	$0{\cdot}4535\rho$†	or	$0{\cdot}6932\rho'$‡	$0{\cdot}2741\rho^2$†	or	$0{\cdot}6404\rho'^2$‡
Normal	$0{\cdot}4431\rho$	or	$0{\cdot}7458\rho'$	$0{\cdot}2357\rho^2$	or	$0{\cdot}6677\rho'^2$
Angular	$0{\cdot}4347\rho$	or	$0{\cdot}8183\rho'$	$0{\cdot}1981\rho^2$	or	$0{\cdot}7020\rho'^2$
Uniform	$0{\cdot}4330\rho$	or	$0{\cdot}9000\rho'$	Not applicable		
One-particle or extreme value	$0{\cdot}4626\rho$	or	$0{\cdot}4126\rho'$	Not applicable		

* The first bound is computed from (5·3) and the second bound from (5·7). (See Table 3 for the values of $\overline{V}$ for the several distributions.)

† σ is the standard deviation of the tolerance distribution and $d = \rho\sigma$.

‡ R is the distance between the 20th and the 80th percentile and $d = \rho' R$.

6. Random location of the dose mesh

In §§ 3–5, the location of the dose mesh, determined by x_0, was assumed to be fixed. If x_0 is assumed to be chosen at random, with distribution uniform on the interval $(0, d)$ say, the estimator, $\bar{x}$, will be unbiased. Denote the expectation over x_0 and over the conditional binomial variables by $E(\bar{x})$. Then

$$E(\bar{x}) = \int_0^d \frac{1}{d} E(\bar{x}|x_0)\,dx_0$$

$$= \frac{1}{d}\int_0^d \sum_{-\infty}^{\infty} \left(x_0 + \frac{d}{2} + id\right) \int_{x_0+id}^{x_0+id+d} dF(x)\,dx_0. \qquad (6{\cdot}1)$$

By integrating first with respect to x_0, then with respect to $F(x)$, and finally summing, the expectation is easily shown to be μ.

If the expectation over x_0 alone is denoted by E_{x_0} and the bias of $\bar{x}$ conditional on x_0 is denoted by $B(\bar{x}|x_0)$, then the mean square error or variance of $\bar{x}$ for random choice of x_0 can be written as follows

$$V(\bar{x}) = E_{x_0}E[(\bar{x}-\mu)^2|x_0], \tag{6·2}$$

$$V(\bar{x}) = E_{x_0}[V(\bar{x}|x_0)] + E_{x_0}[B^2(\bar{x}|x_0)]. \tag{6·3}$$

The first term on the right of (6·3) can be written

$$E_{x_0}[V(\bar{x}|x_0)] = \int_0^d \frac{1}{d}\frac{d^2}{n}\sum_{-\infty}^{\infty} F(x_0+id)\,[1-F(x_0+id)]\,dx_0$$

$$= \frac{d}{n}\int_{-\infty}^{\infty} F(x)\,[1-F(x)]\,dx. \tag{6·4}$$

The final expression in (6·4) is the integral denoted in expression (5·2) by $\bar{V}(\bar{x})$. No simple expression has been found for the second term on the right of (6·3) but it follows from (4·16) that this term is $O(d^2)$. Therefore, it follows that

$$V(\bar{x}) = \bar{V}(\bar{x}) + O(d^2). \tag{6·5}$$

Note that the second term of $V(\bar{x})$ is independent of n and of smaller order in d than the first component, and that $\bar{V}(\bar{x})$ contains d only in the form of the factor d/n, the inverse of the number of subjects tested per unit interval on the dose scale.

The size of the experiment can be measured by the number of subjects tested per unit interval on the dose scale, $n' = n/d$. For fixed n' and random location x_0, the following theorem can be proved.

THEOREM 6·1. $V(\bar{x})$ is a minimum, for fixed n' and random x_0, when $n = 1$ and $d = 1/n'$.

Proof.

$$V(\bar{x}) = \frac{1}{n'}\int_{-\infty}^{\infty} F(1-F)\,dx + \frac{1}{d}\int_0^d B^2(\bar{x}|x_0)\,dx_0. \tag{6·6}$$

The first term on the right of (6·6) is independent of the choice of n, for fixed n'. The proof consists of demonstrating that the second term decreases if n is decreased to 1, so that d becomes $1/n'$. The manipulations involve breaking up the second term in (6·6), for sample size n, into terms corresponding to the smaller intervals of size $1/n'$, examining the difference between biases, conditional on x_0, for the two situations and then integrating out the x_0. The details will not be presented here.

Theorem 6·1 demonstrates that the dose mesh should be as fine as possible for the purpose of estimation by Spearman procedure.

7. ASYMPTOTIC EFFICIENCY OF THE SPEARMAN ESTIMATOR

Suppose that n' increases, and $n = 1$, $d = 1/n'$. Then it follows, for x_0 either fixed or random, that $\bar{x}$ is consistent with a variance that can be written

$$V(\bar{x}) = \frac{1}{n'}\int_{-\infty}^{\infty} F(1-F)\,dx + O\left(\frac{1}{n'^2}\right). \tag{7·1}$$

It is convenient, in the case of a sequence of random variables, to approximate the variances by simpler terms correct to order n^{-1}. (When the sequence does not have variances, the variances of a sequence of limiting distributions are used.) Such a sequence of approximations are called the *asymptotic variances*. In this sense the first term on the right of (7·1) will be called the asymptotic variance of $\bar{x}$ as n' becomes large and will be denoted by $V_A(\bar{x})$:

$$V_A(\bar{x}) = \frac{1}{n'} \int_{-\infty}^{\infty} F(1-F)\,dx. \tag{7·2}$$

Cornfield & Mantel (1950) showed that the Spearman estimator and the maximum likelihood estimator are approximately equal and that the approximation improves without limit as d goes to zero. Bross (1950) and Haley (1952) obtained sampling distributions of the Spearman estimator and the maximum likelihood estimator by enumeration for some specific choices of dose mesh and sample size, for the logistic and normal tolerance distributions respectively. Their results showed that the Spearman estimator was more closely concentrated about the true value, μ, than was the maximum likelihood estimator for every situation considered.

Finney (1950, 1952*a*) computed the mean-square error for the Spearman estimator and the information for the experiment for several choices of x_0, and then averaged over x_0. Enough dose levels were used so that the results were the same as would be obtained by assuming the dose mesh to be infinite. Comparison of the reciprocal of the average information over x_0 with the average square error over x_0 resulted in ratios of 0·9814 for the normal tolerance distribution and 1·0000 for the logistic tolerance distribution.

Finney's criterion for evaluating the Spearman estimator was used here. The concept of information for the finite experiment was extended to the infinite experiment by considering infinite series of information terms corresponding to the doses. The expectation of this series was taken over x_0 with uniform distribution on $(0, d)$. The result is the information (I) for the infinite experiment

$$I = n' \int_{-\infty}^{\infty} \frac{[F_\mu(x)]^2}{F(x)\,[1-F(x)]}\,dx, \tag{7·3}$$

where $F_\mu(x)$ is the partial derivative of the distribution with respect to the single unknown parameter, μ.

The asymptotic efficiency, E, of the Spearman estimator is then defined as the ratio of I^{-1} to the asymptotic variance of the estimator, $V_A(\bar{x})$

$$\begin{aligned} E &= \frac{I^{-1}}{V_A(\bar{x})} \\ &= \left[\int_{-\infty}^{\infty} F(x)\,[1-F(x)]\,dx \int_{-\infty}^{\infty} \frac{F_\mu^2(x)}{F(x)\,[1-F(x)]}\,dx\right]^{-1}. \end{aligned} \tag{7·4}$$

If E is evaluated for the distributions given in Table 1, the following results are obtained: normal, $E = 0{\cdot}9814$; logistic, $E = 1{\cdot}000$; angular, $E = 0{\cdot}8106$; one particle, $E = 0{\cdot}8319$. The definition of information for the uniform distribution is not applicable so the efficiency cannot be computed. Note that the Spearman estimator has full asymptotic efficiency for the infinite experiment when the tolerance distribution is logistic.

The efficiency of the Spearman estimator is not necessarily high. Consider the following

family of symmetrical tolerance distributions with the single unknown parameter, μ, appearing as a translation parameter

$$F_e(x;\mu) = K(e)\int_{-\infty}^{x}\left[1+\frac{(t-\mu)^2}{1+2e}\right]^{-(1+e)}dt. \tag{7·5}$$

It can be shown that the asymptotic variance of the Spearman estimator for this family of tolerance distributions goes to infinity as e goes to zero, while the information is bounded away from zero. Thus the efficiency for this family of tolerance distributions goes to zero as e goes to zero, i.e. as the distributions approach the Cauchy distribution. Since the Spearman estimator simulates an arithmetic mean based on grouped data, this inefficiency is not surprising in view of the known inadequacy of the actual arithmetic mean for the Cauchy distribution itself.

REFERENCES

BERKSON, J. (1955). A statistically precise and relatively simple method of estimating the bioassay with quantal response, based on the logistic function. *J. Amer. Statist. Ass.* **48**, 565–99.

BROSS, I. (1950). Estimates of the $L.D._{50}$: a critique. *Biometrics*, **6**, 413–23.

CORNFIELD, J. & MANTEL, N. (1950). Some new aspects of the application of maximum likelihood to the calculation of the dosage response curve. *J. Amer. Statist. Ass.* **45**, 181–210.

CRAMÉR, H. (1951). *Mathematical Methods of Statistics*. Princeton University Press.

FINNEY, D. J. (1950). The estimation of the mean of a normal tolerance distribution. *Sankhyā*, **10**, 341–60.

FINNEY, D. J. (1952*a*). The estimation of the E.D. 50 for a logistic response curve. *Sankhyā*, **12**, 121–36.

FINNEY, D. J. (1952*b*). *Probit Analysis* (2nd edn.). Cambridge University Press.

HALEY, D. C. (1952). Estimation of the dosage-mortality relationship when the dose is subject to error. *Technical Report*, no. 15. Applied Mathematics and Statistics Laboratory, Stanford University.

SPEARMAN, C. (1908). The method of 'right and wrong cases' ('constant stimuli') without Gauss' formulae. *Brit. J. Psychol.* **2**, 227–42.

Biometrika (1961), **48**, 3 *and* 4, *p.* 303
Printed in Great Britain

Inconsistencies in a schedule of paired comparisons

BY PATRICK SLATER
Institute of Psychiatry, The Maudsley Hospital, University of London

1. THE METHOD, AND THE HYPOTHESES CONCERNING IT

The method of paired comparison has had a long and honourable history in psychological experiments, beginning with the researches of Witmer and Cohn, published in 1894. Titchener (1901) described it in detail in one of the earliest text-books on experimental psychology and Guilford (1954) devotes a chapter to it in the latest edition of his popular text-book. Theoretical investigations of the method, which has applications outside psychology, still continue to appear, e.g. by David (1959) in this Journal, in technical reports by Gulliksen & Tucker (1959) and in a thesis by the author (1960). The authoritative paper on the null hypothesis concerning it is the one by Kendall & Babington Smith which appeared here in 1939. With this I find myself in disagreement.

The experimental procedure is to show a set of m objects to an individual in pairs and ask him each time to choose one. It is always understood that the objects differ from one another, but there may be doubt whether the difference is discernible by the individual. The difference may be confined to one respect, e.g. a set of boxes may be used identical in appearance but differing in weight, and the observer's attention may be directed to that respect, e.g. by the instruction, 'Choose the heavier each time'. Or they may differ in several respects and the criterion of choice may be left to the individual, e.g. in Titchener's standard procedure the objects are coloured cards differing in hue and saturation and the individual is instructed to choose whichever he prefers. It is normally understood, but not always, cf. Myers (1925), that each of the $\frac{1}{2}m(m-1)$ possible pairs is presented once and once only.

We shall assume that the objects may differ in several respects and that the individual's attention has not been directed to any respect for which there is an independent criterion; also that he has been shown every possible pair once and is never permitted to evade the obligation to choose, e.g. by responding, 'Both alike'. The objects will be denoted $A, B, \ldots, M$.

Initially we may hope to show that the individual is aware of *one* dimension of preference, in accordance with which the objects can be arranged in an order from most preferred to least. The contrary, C_1, which must be disproved before any such hypothesis, H_1, need be conceded, is that the individual is unaware of any differences between the objects and that all his choices are made at random, independently of one another. It may be disproved if an unexpectedly large number of the choices are internally consistent, i.e. cohere with the same one out of all the $m!$ possible orders for m objects; for in the absence of any criterion all possible orders are equally admissible. The minimum number of inconsistent responses will be denoted by i, and an order with which there are only i inconsistent responses will be called a nearest adjoining order.

In some specimen schedules of responses the nearest adjoining order is not unique; there may be several orders, say j altogether, with only i inconsistencies. The numbers

i and j are always ascertainable, at least in theory, if the schedule is checked against the $m!$ possible orders. Consider, for instance, $A > B$, $A < C$, $B > C$, a specimen schedule for 3 objects (read $>$ as 'preferred to'). One of the responses is inconsistent with each of the three orders $A > B > C$, $B > C > A$ and $C > A > B$. The other possible orders, $A > C > B$, $B > A > C$ and $C > B > A$, can be omitted from consideration for two responses are inconsistent with each of them. The nearest adjoining order is not unique so the inconsistent response cannot be identified, but certainly $i = 1$ and $j = 3$.

The sample space or universe for $m = 3$ contains 8 distinct specimens of possible schedules of responses; and all are equiprobable on C_1. It is easy to verify that six have $i = 0$ and two have $i = 1$. So a schedule with $i = 0$ is not exceptional when $m = 3$ and C_1 is tenable for all the specimens in this universe. In general the universe for m objects contains $2^{\binom{m}{2}}$ equiprobable specimens. Let s_c be any specimen with a certain number of inconsistencies, i.e. with $i = c$, and let $f_m(c)$ be the frequency of occurrence of all such specimens in the universe for m. The question to be decided is whether C_1 is tenable concerning a particular s_c from this universe. If we make it our rule to reject C_1 when its probability is below 0·05 we can reach a decision if we know what is the limiting values of i, say $i = u$, for which

$$\sum_{i=0}^{u} f_m(i)/2^{\binom{m}{2}} < 0{\cdot}05.$$

Then we reject C_1 if $c \leqslant u$ but not otherwise.

Consider next the s_c with $c \leqslant u$ in different universes, for all of which H_1 must be admitted. As m and consequently u are allowed to increase, c can increase indefinitely without exceeding u. At some point it may begin to seem surprising that so many responses, which can be itemized if the nearest adjoining order is unique, are all consistent with one another, i.e. with one other ordering of the objects, which may be called the residual order; and we may feel tempted to consider the more elaborate hypothesis, H_2, that the individual is aware of *two* dimensions of preference. For the objects can be arrayed on a surface definable by two axes, falling in the nearest adjoining order along one and in the residual order along the other, so that every choice appears consistent with one or other of the two orders. The contrary we now encounter, C_2, which must be disproved before H_2 need be conceded, is that the choices inconsistent with the nearest adjoining order do not imply any awareness of a second dimension but are all made at random and independently of one another. The argument of § 3 below is that under the conditions of the experiment C_2 is always tenable, no matter how large m and i are.

The hypotheses under consideration all relate to the $\binom{m}{2}$ responses the individual is required to make. Each is the result of a single act of choice, potentially independent of every other such act and liable to bring the laws of probability into operation. So I regard the responses as the simple events from which the universe for m originates and conclude that the probability distribution for i, the number of inconsistent responses, is what needs to be examined when C_1 is under consideration—not the probability distribution for d, the number of circular triads in a schedule, which is the variable considered by Kendall & Babington Smith.

A triad is the set of responses relating to three objects. It may have $i = 0$ or 1, as already mentioned, and it is circular when $i = 1$. The authors only give reasons of simplicity and convenience for treating triads as units for enumeration. After describing them and larger

polyads within the complete configuration or m-ad which may be used to represent the schedule of responses, they remark 'it seems best to confine attention to circular triads, which, so to speak, constitute the inconsistent elements in the configuration, and to ignore the more ambiguous criteria associated with circular polyads of greater extent'. They do not mention the possibility of treating each response as a unit, nor do they offer any explicit definition of the null hypothesis to be considered.

Triads ought not to be treated as elements. They are compound events not conceivably independent of one another, for the total number of triads in a schedule exceeds the number of responses by a factor of $\frac{1}{3}(m-2)$ and each response features in $m-2$ triads. Moreover, there is no 1 : 1 relationship between i and d; schedules from the same m with the same i may differ in d, and vice versa. For instance,

$$\text{when } i = 1, \quad d \text{ ranges from 1 to } m-2,$$

$$\text{when } i = 2, \quad d \text{ ranges from 2 to } 2m-6;$$

and further evidence appears in Fig. 1 and Table 2. If the inconsistencies were subclassified d might be defined as a weighted summation of the frequencies in specified classes: that is to say, d may be viewed as a summation in which some inconsistencies receive more weight than others. But on the assumption that the inconsistent responses result from erratic acts of choice and occur at random there is no justification for subclassifying them. And the conditions of the experiment do not include any region where this assumption can be proved to have a negligible probability.

Inconsistent responses receive equal weights in Kendall's procedure for τ (1938, 1948), so there is a simple relationship between τ and i. A nearest adjoining order might be defined as any order which maximizes τ for the schedule under consideration, and the maximum value of τ is obtainable from i, given m, as $1-4i/m(m-1)$.

2. The first form of the null hypothesis, C_1

On C_1 when two of the objects, I and J, are presented to the individual, since he is unaware of any difference between them but obliged to make a choice, he is just as likely to choose I or J, and his choice will be not influenced by any choices he may have made on any previous occasion when he may have had I or J presented for comparison with any of the other objects.

The universe of different schedules of responses thus obtainable consists of $2^{\binom{m}{2}}$ specimens, all equiprobable. This total needs to be broken down into subtotals for specimens where $i = 0, 1, 2, \ldots$. Then to decide whether C_1 applies to a particular schedule for m objects we need to find the number of inconsistencies in it and see what proportion of the schedules in the universe contains no more than the same number of inconsistencies.

Table 1 shows the breakdowns for $m \leqslant 8$, and the cumulative proportions derived. If C_1 is considered acceptable at the 0·05 probability level but not below, it is tenable for all schedules where $m < 6$, but not when $i = 0, m \geqslant 6$, when $i = 1, m \geqslant 7$, or when $i = 2, m \geqslant 8$.

The frequency distribution for any m may be defined as the expansion

$$2^{\binom{m}{2}} = f_m(0) + f_m(1) + f_m(2) + \ldots.$$

A general algebraic definition of $f_m(i)$ would thus define the complete frequency

distribution for all m. Such an expression has not yet been found, but the expressions for values of i up to 3 can be given. They are

$$f_m(0) = m!,$$

$$f_m(1) = m!\,(3m^2 - 13m + 14)/6,$$

$$f_m(2) = m!\,(9m^4 - 78m^3 + 235m^2 - 438m + 680)/72,$$

$$f_m(3) = m!\,(135m^6 - 1{,}755m^5 + 8{,}685m^4 - 27{,}185m^3 + 77{,}820m^2 - 157{,}204m + 210{,}336)/6480.$$

The expression for $f_m(2)$ only applies when $m \geqslant 4$, and the expression for $f_m(3)$ only when $m \geqslant 6$. When $m = 5$, $f_5(3) = 24$.

These expressions give $\sum_{i=0}^{3} f_m(i)/2^{\binom{m}{2}} = 0{\cdot}009902$

for $m = 9$. So u is certainly not less than 3 when $m \geqslant 9$. It may even exceed 3, and appears to be increasing at an accelerating rate.

Table 1. *The frequency distribution of i for given values of m*

i	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$
				Part 1			
0	2	6	24	120	720	5,040	40,320
1	—	2	40	480	5,280	58,800	685,440
2	—	—	—	400	13,280	278,880	5,120,640
3	—	—	—	24	11,568	651,504	21,590,016
4	—	—	—	—	1,920	736,848	55,101,312
5	—	—	—	—	—	323,120	84,325,248
6	—	—	—	—	—	41,040	71,687,040
7	—	—	—	—	—	1,920	27,421,440
8	—	—	—	—	—	—	2,464,000
				Cumulative proportions			
				Part 2			
0	1·0	0·750	0·375	0·11719	0·02197	0·002403	0·000150
1	—	1·0	1·0	0·58594	0·18311	0·030441	0·002704
2	—	—	—	0·97656	0·58838	0·163422	0·021780
3	—	—	—	1·0	0·94141	0·474083	0·102209
4	—	—	—	—	1·0	0·825439	0·307477
5	—	—	—	—	—	0·979515	0·621613
6	—	—	—	—	—	0·999084	0·888668
7	—	—	—	—	—	1·0	0·990821
8	—	—	—	—	—	—	1·0

The data for $m \geqslant 6$ have been provided by the National Physical Laboratory using an electronic computing programme developed by G. G. Alway as a research project. An account of it will be published separately. Considerable expense would be incurred if the research were continued to obtain complete expansions of $2^{\binom{m}{2}}$ for larger m or expressions defining $f_m(i)$ for larger i, so that it seems desirable to publish the present results and to ascertain the consensus of expert opinion before proceeding. For most practical purposes it would be sufficient to know the values of u for $m \leqslant 15$. It is true that experiments have

been conducted with considerably more than 15 objects. Titchener (1901) regularly used one with $m = 27$ for his course in experimental psychology and Cattell, Maxwell, Light & Unger (1949) have described one with $m = 50$. Experiments with large m need not be difficult to conduct if the objects are suitably chosen and appropriate apparatus is constructed for presenting them in pairs and recording the responses automatically, but it is not often that any compelling reasons for conducting such experiments are encountered in practice.

3. The second form, C_2; and an argument that every inconsistency should be given an equal weight

The evidence of a single schedule of responses is never sufficient to make C_2 untenable. Take $A > B > \ldots > M$ arbitrarily as an order with which some of the individual's responses cohere. Then the remainder must all cohere with the opposite order $A < B < \ldots < M$. If none are consistent with the first all are consistent with the second. So the i responses inconsistent with the nearest adjoining order must *a fortiori* all be consistent with one another, and in general may be linked together in many different ways to form possible residual orders. In other words evidence in favour of H_2 is indistinguishable from evidence against it, so C_2 cannot be ruled out. If we wish to disprove C_2 we must adduce supplementary evidence from other sources, modify the conditions of the experiment in some way or advance some specific argument.

Thus the only alternatives to be considered when investigating the internal consistency of a single schedule of responses are

(i) C_1: the observer is unable to discriminate between the objects, or

(ii) H_1+C_2: he is aware of a single dimension of preference. Choices not made in accordance with it are produced by chance causes, i.e. causes operating independently on particular judgements, such as distractions or momentary lapses of attention, etc.

There is never any case for pressing on to consider alternatives such as might be denoted $H_1+H_2+C_3$, etc., without additional evidence.

We may argue from this that every inconsistency should be given an equal weight when C_1 is under consideration. If the order $A > B > \ldots > M$ is the dimension of preference characteristic of the individual, a cause operating accidentally is just as likely to produce the reversal $A < M$ as $A < D$, say, and as no causes other than accidental causes need be supposed, we ought not to assign more weight to one such reversal than another. A straight count, that is to say, an unweighted summation of the inconsistencies is therefore the index we should use in deciding whether H_1 or C_1 is to be preferred.

The proposition can be sustained, perhaps quite adequately, without reference to C_2. For *per contra* we cannot claim that $A < M$ should be given a greater weight than $A < D$ without postulating that A is further removed from M than from D on the scale of preference characteristic of the individual. But this is to concede a form of H_1, and we should not make any such concession before we have succeeded in disproving C_1. Moreover, the relative weights we assign to $A < D$ and $A < M$ must depend on the particular form of H_1 we choose to concede; but even after disposing of C_1 we may be left with $j > 1$, i.e. with several equally acceptable forms of H_1.

4. COUNTING i IN A SCHEDULE OF PAIRED COMPARISONS OBTAINED EXPERIMENTALLY

Mr Alway has kindly contributed the following practical notes:

No simple rule for obtaining i is known to be applicable in all cases, but in all practical applications encountered so far the following simple rules have sufficed.

First re-order the rows and columns of the preference matrix* so that the numbers of $+$'s in successive rows are in descending order. Next, examine each row to the right of the diagonal element, and proceeding element by element count separately the positive and negative ones. If at any stage the number of negative elements exceeds the number of positive ones the matrix may be transformed to decrease to total number of negative elements in the upper triangular half. For example, if one of the rows (starting with the diagonal element) is

$$. + + - + - - + - - + + +$$

then by placing this row and the corresponding column nine places further on, the total number of negative elements in the upper half is reduced by 1. The resultant matrix can be examined again in this way. The columns should also be examined in a corresponding fashion; this is the same as examining the rows starting with the diagonal element and counting backwards towards the first element. The process should then be repeated, and wherever the count of $-$'s equals the count of $+$'s the matrix should be transformed in a corresponding fashion. This change will not of itself reduce the total number of $-$'s, but it may alter the position of certain elements so that the number of $-$'s in some row or column exceeds the count of $+$'s, and then the total number may be reduced. This process has sufficed in all practical applications ($m \leqslant 10$) to reduce the number of inconsistencies to its lowest value, and also to give all the permutations for which this lowest value is attained.

Even when the first rule, to put the number of $+$'s in successive rows in descending order, is omitted, the simplest case in which the second rule, of counting $+$'s and $-$'s by row and column, fails by itself to give i is

$$\begin{array}{cccccccc}
. & + & + & - & + & - & + & + \\
- & . & + & + & + & + & + & + \\
- & - & . & + & + & - & + & - \\
+ & - & - & . & + & + & + & + \\
- & - & - & - & . & + & + & - \\
+ & - & + & - & - & . & + & + \\
- & - & - & - & - & - & . & + \\
- & - & + & - & + & - & - & .
\end{array}$$

for which $i = 4$, given by the permutation (24618357) of the rows and columns.

The calculation of the data mentioned in § 2 and the writing of this section have been carried out as part of the research programme of the National Physical Laboratory and they are published by permission of the Director of the Laboratory.

* A preference matrix for recording the responses in a schedule is an $m \times m$ table with a row and corresponding column for each object. The response $I > J$ is recorded as $+1$, or simply $+$, in row I column J, and as -1, or simply $-$, in row J column I.

5. Kendall's d and its relation to i

When one of three conjoined choices is inconsistent with the other two the triad formed is circular. Kendall & Babington Smith's procedure depends on finding the number, d, of such triads in a schedule. The simple computing method for finding d is a great advantage of their procedure. Counting the number of + 's in each row of the preference matrix **A** provides a column vector **a** which is a partition of $\frac{1}{2}m(m-1)$; and d can be obtained from

$$2d = m(m-1)(2m-1)/6 - \mathbf{a}'\mathbf{a}.$$

$\mathbf{A}_1$					$\mathbf{a}_1$
.	+	−	+	+	3
−	.	+	+	+	3
+	−	.	+	+	3
−	−	−	.	+	1
−	−	−	−	.	0

$\mathbf{A}_2$					$\mathbf{a}_2$
.	+	+	+	−	3
−	.	+	+	+	3
−	−	.	+	+	2
−	−	−	.	+	1
+	−	−	−	.	1

$\mathbf{A}_3$					$\mathbf{a}_3$
.	+	−	+	+	3
−	.	+	+	+	3
+	−	.	+	−	2
−	−	−	.	+	1
−	−	+	−	.	1

Fig. 1. Three preference matrices, and their partitions.

For example, the three matrices in Fig. 1 with partitions as shown have

	$\mathbf{a}'\mathbf{a}$	d	i	j
$\mathbf{A}_1$	28	1	1	3
$\mathbf{A}_2$	24	3	1	1
$\mathbf{A}_3$	24	3	2	5

It appears debatable whether the inconsistency in $\mathbf{A}_1$ should be given more or less weight than the one in $\mathbf{A}_2$. For instance, it might be argued from j that the one in $\mathbf{A}_2$ is the more reasonably attributable to some accidental cause, as it does not evoke any doubt about what order represents the individual's characteristic dimension of preference. My view, based on the argument in § 3, is that both inconsistencies should receive the same weight. Kendall's procedure weights the inconsistency in $\mathbf{A}_2$ three times as heavily as the one in $\mathbf{A}_1$.

Moreover, Fig. 1 shows that no simple relationship exists between i and d: $\mathbf{A}_1$ and $\mathbf{A}_2$ have the same i but a different d, $\mathbf{A}_2$ and $\mathbf{A}_3$ have the same d but a different i. Table 2 shows the relationship between i and d in the universes for $m \leqslant 8$. The two quantities are quite closely correlated, viz.

In the universe for	r is
$m = 4$	0·9317
5	·9087
6	·9031
7	·8969
8	·8927

It is not surprising that r diminishes as m increases. Increasing m provides more freedom for preference matrices with the same partition to vary in the internal arrangements for their + 's and − 's.

The correlation is not close enough to prevent different results being obtained when d and i are used to test C_1 with reference to a single schedule of responses. When $m = 7$ d leads to the rejection of some **A**'s at the 5 % significance level where $i = 1$ and the acceptance of others where $i = 2$; and when $m = 8$ to the rejection of some where $i = 2$ and the acceptance of others where $i = 2$ or even 4. The 5·85 million possible schedules acceptable

Table 2. *The association between i and d shown by the frequencies of schedules with given i and d in the universes for m = 4, 5, 6, 7 and 8*

Case m = 4

d \ i	0	1	Total
2	—	24	24
1	—	16	16
0	24	—	24
Total	24	40	64

Case m = 5

d \ i	0	1	2	3	Total
5	—	—	—	24	24
4	—	—	280	—	280
3	—	120	120	—	240
2	—	240	—	—	240
1	—	120	—	—	120
0	120	—	—	—	120
Total	120	480	400	24	1,024

Case m = 6

d \ i	0	1	2	3	4	Total
8	—	—	—	720	1,920	2,640
7	—	—	—	4,800	—	4,800
6	—	—	2,880	5,760	—	8,640
5	—	—	3,360	288	—	3,648
4	—	720	5,520	—	—	6,240
3	—	1,440	1,440	—	—	2,880
2	—	2,160	80	—	—	2,240
1	—	960	—	—	—	960
0	720	—	—	—	—	720
Total	720	5,280	13,280	11,568	1,920	32,768

Case m = 7

d \ i	0	1	2	3	4	5	6	7	Total
14	—	—	—	—	—	—	720	1,920	2,640
13	—	—	—	—	—	31,920	40,320	—	72,240
12	—	—	—	—	52,080	181,440	—	—	233,520
11	—	—	—	5,040	139,440	99,680	—	—	244,160
10	—	—	—	50,400	323,568	10,080	—	—	384,048
9	—	—	—	131,040	149,520	—	—	—	280,560
8	—	—	20,160	184,800	72,240	—	—	—	277,200
7	—	—	30,240	157,920	—	—	—	—	188,160
6	—	—	70,560	109,200	—	—	—	—	179,760
5	—	5,040	72,240	13,104	—	—	—	—	90,384
4	—	10,080	65,520	—	—	—	—	—	75,600
3	—	15,120	18,480	—	—	—	—	—	33,600
2	—	20,160	1,680	—	—	—	—	—	21,840
1	—	8,400	—	—	—	—	—	—	8,400
0	5,040	—	—	—	—	—	—	—	5,040
Total	5,040	58,800	278,880	651,504	736,848	323,120	41,040	1,920	2,097,152

Table 2 (*cont.*)

Case $m = 8$

d \ i	0	1	2	3	4	5	6	7	8	Total
20	—	—	—	—	—	—	40,320	1,612,800	1,576,960	3,230,080
19	—	—	—	—	—	—	1,236,480	8,279,040	887,040	10,402,560
18	—	—	—	—	—	483,840	11,544,960	12,284,160	—	24,312,960
17	—	—	—	—	—	3,010,560	17,310,720	4,166,400	—	24,487,680
16	—	—	—	—	483,840	11,329,920	24,326,400	1,048,320	—	37,188,480
15	—	—	—	—	1,317,120	16,251,648	11,719,680	—	—	29,288,448
14	—	—	—	40,320	5,362,560	24,465,280	4,863,360	30,720	—	34,762,240
13	—	—	—	403,200	8,682,240	14,784,000	645,120	—	—	24,514,560
12	—	—	—	1,290,240	12,566,400	10,631,040	—	—	—	24,487,680
11	—	—	—	1,908,480	9,639,168	3,207,680	—	—	—	14,755,328
10	—	—	161,280	4,179,840	11,319,168	161,280	—	—	—	15,821,568
9	—	—	241,920	4,058,880	4,032,000	—	—	—	—	8,322,800
8	—	—	645,120	4,587,520	1,693,440	—	—	—	—	6,926,080
7	—	—	779,520	3,091,200	—	—	—	—	—	3,870,720
6	—	40,320	1,249,920	1,747,200	5,376	—	—	—	—	3,042,816
5	—	80,640	967,680	256,256	—	—	—	—	—	1,304,576
4	—	120,960	806,400	26,880	—	—	—	—	—	954,240
3	—	161,280	241,920	—	—	—	—	—	—	403,200
2	—	201,600	26,880	—	—	—	—	—	—	228,480
1	—	80,640	—	—	—	—	—	—	—	80,640
0	40,320	—	—	—	—	—	—	—	—	40,320
Total	40,320	685,440	5,120,640	21,590,016	55,101,312	84,325,248	71,697,040	27,421,440	2,464,000	268,435,456

in accordance with i and the 9·92 million acceptable in accordance with d when $m = 8$ include 4·80 million in common. There is disagreement about the remainder. Part of this disagreement arises because i and d are discrete variables, so that the tails of their probability distributions cannot be cut off exactly at the 0·05 level. In percentages, 2·18 pass the test on i, 3·70 pass on d, 1·79 pass on both.

When several individuals, say n altogether, are asked to compare the same m objects in pairs and little evidence of agreement is found between them, the question may arise whether the absence of agreement reflects differences in taste or lack of discernment. It should be possible to extend the use of i or d to consider problems of this kind, and the correlation between them should be sufficient to lead to convergent conclusions when n is not too small. For n above a certain limit the advantage of easy computation might tell decisively in favour of d.

I would like to emphasize the importance of distinguishing between problems of discernment and problems of agreement in this context. Comparison in pairs is specially appropriate for problems of discernment; it provides more evidence of internal consistency, or the lack of it, than comparison in sets of more than two at a time. But for investigating problems of agreement it does not appear to have any advantages over other methods of multiple comparison, of which ranking is administratively the most convenient.

REFERENCES

CATTELL, R. B., MAXWELL, E. F., LIGHT, B. H. & UNGER, M. P. (1949). The objective measurement of attitudes. *Brit. J. Psychol.* **40**, 81–90.

COHN, JONAS (1894). Experimentelle Untersuchungen über die Gefühlsbetonungen der Farben, Helligkeiten und ihrer Combinationen. *Philos. Stud. Leipz.* **10**, 562–603.

DAVID, H. A. (1959). Tournaments and paired comparisons. *Biometrika*, **46**, 139–49.

GUILFORD, J. P. (1954). *Psychometric Methods.* New York: McGraw Hill Book Co. Inc.

GULLIKSEN, H. & TUCKER, L. R. (1959). A general procedure for obtaining paired comparisons from multiple rank orders. Princeton University, NR 150–088.

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81–93.

KENDALL, M. G. (1948). *Rank Correlation Methods.* London: Charles Griffin and Co. Ltd.

KENDALL, M. G. & BABINGTON SMITH, B. (1939). On the method of paired comparisons. *Biometrika*, **31**, 324–45.

MYERS, C. S. (1925). *A Text Book of Experimental Psychology.* Cambridge University Press.

SLATER, P. (1960). The reliability of some methods of multiple comparison in psychological experiments. London University thesis.

TITCHENER, E. B. (1901). *Experimental Psychology. I. Qualitative.* London: Macmillan and Co. Ltd.

WITMER, L. (1894). Zur experimentellen Aesthetick einfacher räumliche Formverhältnisse. *Philos. Stud. Leipz.* **9**, 96–144 and 209–63.

Biometrika (1961), **48**, 3 *and* 4, *p.* 313
With 4 *text-figures*
Printed in Great Britain

Some circular coverage problems†

BY GEORGE W. MORGENTHALER
The Martin Co., Denver, Colorado

1. THREE COVERAGE PROBLEMS

The following three coverage problems are of mathematical and military interest.

Problem 1. A damage circle of radius R is aimed at the centre, $(0, 0)$, of a circular target region of radius Z (Fig. 1). What is the expected or average coverage of the target circle if the impact points (x, y) are circularly normally distributed, with standard deviation σ?

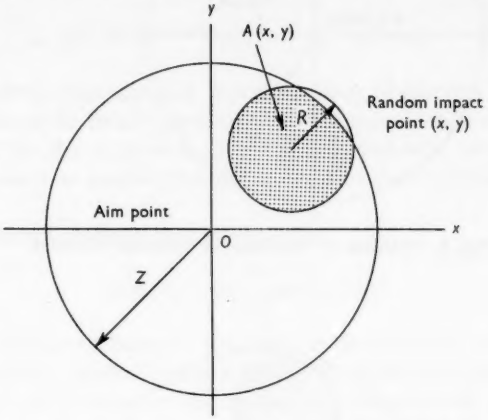

Fig. 1. Problem 1. Coverage by one aimed circle.

Problem 2. A cluster of n circles is aimed at $O = (0, 0)$, the centre of the target circle of radius Z (Fig. 2). If the cluster circles are non-overlapping—each of radius R—and if the centre of the cluster is circularly normally distributed about O with standard deviation σ, what is the expected coverage?

Problem 3. A set of n independently aimed circles are impacted at the target circle of radius Z (Fig. 3). The ith of the impacting circles is of radius R and is aimed at an aim point $A_i = (a_i, b_i)$ $(i = 1, 2, 3, \ldots, n)$, each impact being circularly normally distributed about its aim point with standard deviation σ. What is the expected coverage?

These three problems will be considered in the section below, which will include a summary of the rather scattered work of a number of authors and also, it is believed, some new results under the following headings:

(*a*) Introduction of the diffused exponential bomb (covering circle) as a means of arriving at approximate solutions to the offset circle *v.* circle coverage problem.

† A partial summary of this paper was presented at the Seventeenth Annual meeting of the Operations Research Society of America in New York, 21 May 1960. The paper is based on Martin Company Report P-59-69.

(*b*) Use of the above offset circle *v.* circle diffused bomb model to solve the cluster-of-circles coverage problem.

(*c*) Use of the above offset circle *v.* circle diffused bomb model to solve the problem of several individually aimed circles *v.* a target circle.

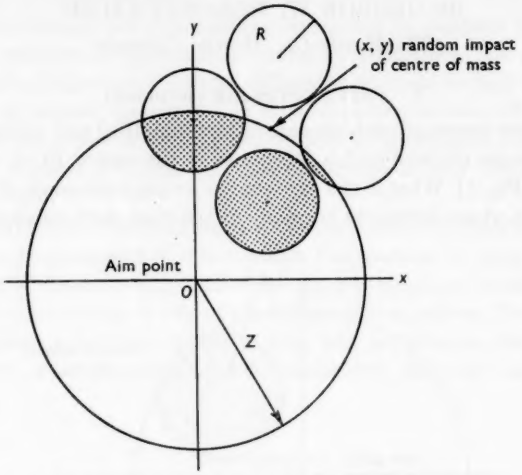

Fig. 2. Problem 2. Coverage by a cluster of circles.

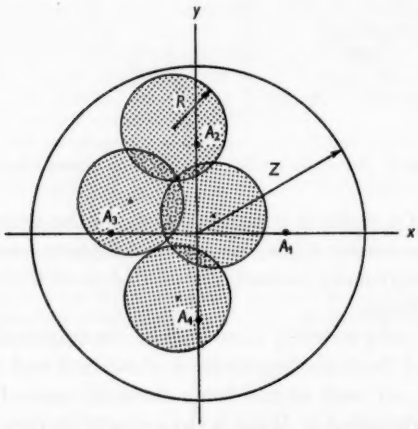

Fig. 3. Problem 3. Coverage by independently aimed circles.

2. EXPONENTIAL MODEL FOR EXPECTED COVERAGE WITH SINGLE CIRCLE

Using the 'cookie-cutter' or circle of damage approach, the expected damage, E, may be written as

$$E = \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(x,y)\, e^{-\frac{1}{2}(x^2+y^2)/\sigma^2}\, dx\, dy, \tag{1}$$

where $A(x,y)$ is the overlap when impact is at (x,y) (see shaded region in Fig. 1).

In Report P-59-69 of The Martin Co., the author has traced the interesting attempts to solve this problem at the Rand Corporation and elsewhere. These include approximations to $A(x, y)$, use of squares or rectangles in place of circles, and special analogue integrating devices. Germond (1950) has provided a solution in terms of the so-called offset circular probability function, as follows. Let $a^2+b^2 = r^2$, and

$$p(R,r) = \frac{1}{2\pi}\iint_{(x-a)^2+(y-b)^2\leqslant R^2} e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy$$

$$= \frac{1}{2\pi}\iint_{\xi^2+\eta^2\leqslant R^2} \exp\left[-\tfrac{1}{2}\{(\xi+a)^2+(\eta+b)^2\}\right]d\xi\,d\eta. \tag{2}$$

Then $p(R, r)$ is the probability that a random point circularly normally distributed about the origin with unit variance will fall within a circle of radius R about point (a, b). This function and its complement, $q(R, r)$, are called offset circular functions. $q(R, r)$ can be shown to be equal to

$$q(R,r) = e^{-\frac{1}{2}R^2} + R\int_0^r e^{-\frac{1}{2}(R^2+t^2)} I_1(Rt)\,dt, \tag{3}$$

where $I_1(Rt)$ is the first-order modified Bessel function of the first kind. The National Bureau of Standards and the Rand Corporation have prepared tables of the function $q(R, r)$ in increments of 0·1 in the arguments (The Rand Corporation, 1952). Weil (1954) and Laurent (1957) also describe methods for finding the probability of being within a distance R of (a, b).

Using $p(R, r)$, Germond observed that the function

$$S(Z,R) = 2\pi\int_0^R \rho p(Z,\rho)\,d\rho, \tag{4}$$

which provides the solution of Problem 1, is equal to the expected overlap when a circle of radius R is dropped over a circle of radius Z if the point of aim is the centre of the stationary circle and if the impact distribution for the centre of the damage circle is circularly normally distributed with unit standard deviation. Germond expressed $S(Z, R)$ in terms of the offset circle function $p(Z, R)$ and the Bessel functions $I_0(ZR)$ and $I_1(ZR)$. A table of the relative target coverage, $S(Z, R)/(\pi Z^2)$, is found in Germond (1950, pp. 15, 16). The table is in jumps of 0·5 at first in Z and then in unit jumps, and in 0·1 increments of R.

Von Neumann, Thompson (1958), and others utilized a 'diffused target' idea with success in such problems. Here we shall diffuse the damage function. Instead of

$$P\{\text{damage of point } (\xi,\eta)|\text{impact at } (x,y)\} = \begin{cases} 1 & \text{if } (x-\xi)^2+(y-\eta)^2 \leqslant R^2, \\ 0 & \text{otherwise,} \end{cases}$$

define $$P\{\text{damage of point } (\xi,\eta)|\text{impact at } (x,y)\} = \exp\left(-\frac{(\xi-x)^2+(\eta-y)^2}{2B^2}\right), \tag{5}$$

B being a constant.

The probability of damage to point (ξ, η), given that (x, y) is circularly normally distributed about the origin with variance σ^2, is then

$$f(\xi,\eta) = \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{(\xi-x)^2+(\eta-y)^2}{2B^2}\right)\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)dx\,dy$$

$$= \frac{B^2}{\sigma^2+B^2}\exp\left(-\frac{\xi^2+\eta^2}{2(\sigma^2+B^2)}\right). \tag{6}$$

Table 1. *B/Z as a function of R/Z*

$\frac{Z}{B\sqrt{2}}$	$\frac{R}{Z}$	$\frac{B}{Z}$	$\frac{Z}{B\sqrt{2}}$	$\frac{R}{Z}$	$\frac{B}{Z}$	$\frac{Z}{B\sqrt{2}}$	$\frac{R}{Z}$	$\frac{B}{Z}$	$\frac{Z}{B\sqrt{2}}$	$\frac{R}{Z}$	$\frac{B}{Z}$
0·05	0·9994	14·1421	2·55	0·3919	0·2773	5·05	0·1980	0·1400	7·55	0·1325	0·0937
·10	·9975	7·0711	2·60	·3844	·2720	5·10	·1961	·1386	7·60	·1316	·0930
·15	·9944	4·7140	2·65	·3772	·2668	5·15	·1942	·1373	7·65	·1307	·0924
·20	·9901	3·5355	2·70	·3702	·2619	5·20	·1923	·1360	7·70	·1299	·0918
·25	·9846	2·8284	2·75	·3635	·2571	5·25	·1905	·1347	7·75	·1290	·0912
0·30	0·9779	2·3570	2·80	0·3571	0·2525	5·30	0·1887	0·1334	7·80	0·1282	0·0907
·35	·9701	2·0203	2·85	·3508	·2481	5·35	·1869	·1322	7·85	·1274	·0901
·40	·9613	1·7678	2·90	·3448	·2438	5·40	·1852	·1309	7·90	·1266	·0895
·45	·9514	1·5713	2·95	·3390	·2397	5·45	·1835	·1297	7·95	·1258	·0889
·50	·9406	1·4142	3·00	·3333	·2357	5·50	·1818	·1286	8·00	·1250	·0884
0·55	0·9289	1·2856	3·05	0·3279	0·2318	5·55	0·1802	0·1274	8·05	0·1242	0·0878
·60	·9164	1·1785	3·10	·3226	·2281	5·60	·1786	·1263	8·10	·1235	·0873
·65	·9031	1·0879	3·15	·3175	·2245	5·65	·1770	·1252	8·15	·1227	·0868
·70	·8891	1·0102	3·20	·3125	·2210	5·70	·1754	·1241	8·20	·1220	·0862
·75	·8745	·9428	3·25	·3077	·2176	5·75	·1739	·1230	8·25	·1212	·0857
0·80	0·8594	0·8839	3·30	0·3030	0·2143	5·80	0·1724	0·1219	8·30	0·1205	0·0852
·85	·8438	·8319	3·35	·2985	·2111	5·85	·1709	·1209	8·35	·1198	·0847
·90	·8279	·7857	3·40	·2941	·2080	5·90	·1695	·1198	8·40	·1190	·0842
·95	·8116	·7443	3·45	·2899	·2050	5·95	·1681	·1188	8·45	·1183	·0837
1·00	·7951	·7071	3·50	·2857	·2020	6·00	·1667	·1179	8·50	·1176	·0832
1·05	0·7784	0·6734	3·55	0·2817	0·1992	6·05	0·1653	0·1169	8·55	0·1170	0·0827
1·10	·7616	·6428	3·60	·2778	·1964	6·10	·1639	·1159	8·60	·1163	·0822
1·15	·7448	·6149	3·65	·2740	·1937	6·15	·1626	·1150	8·65	·1156	·0817
1·20	·7280	·5893	3·70	·2703	·1911	6·20	·1613	·1140	8·70	·1149	·0813
1·25	·7112	·5657	3·75	·2667	·1886	6·25	·1600	·1131	8·75	·1143	·0808
1·30	0·6946	0·5439	3·80	0·2632	0·1861	6·30	0·1587	0·1122	8·80	0·1136	0·0804
1·35	·6782	·5238	3·85	·2597	·1837	6·35	·1575	·1114	8·85	·1130	·0799
1·40	·6621	·5051	3·90	·2564	·1813	6·40	·1563	·1105	8·90	·1124	·0795
1·45	·6462	·4877	3·95	·2532	·1790	6·45	·1550	·1096	8·95	·1117	·0790
1·50	·6306	·4714	4·00	·2500	·1768	6·50	·1538	·1088	9·00	·1111	·0786
1·55	0·6153	0·4562	4·05	0·2469	0·1746	6·55	0·1527	0 1080	9·05	0·1105	0·0781
1·60	·6004	·4419	4·10	·2439	·1725	6·60	·1515	·1071	9·10	·1099	·0777
1·65	·5858	·4285	4·15	·2410	·1704	6·65	·1504	·1063	9·15	·1093	·0773
1·70	·5717	·4159	4·20	·2381	·1684	6·70	·1493	·1055	9·20	·1087	·0769
1·75	·5579	·4041	4·25	·2353	·1664	6·75	·1481	·1048	9·25	·1081	·0764
1·80	0·5446	0·3928	4·30	0·2326	0·1644	6·80	0·1471	0·1040	9·30	0·1075	0·0760
1·85	·5316	·3822	4·35	·2299	·1626	6·85	·1460	·1032	9·35	·1070	·0756
1·90	·5191	·3722	4·40	·2273	·1607	6·90	·1449	·1025	9·50	·1053	·0744
1·95	·5071	·3626	4·45	·2247	·1589	6·95	·1439	·1017	10·00	·1000	·0707
2·00	·4954	·3536	4·50	·2222	·1571	7·00	·1429	·1010	12·00	·0833	·0589
2·05	0·4841	0·3449	4·55	0·2198	0·1554	7·05	0·1418	0·1003	15·00	0·0667	0·0471
2·10	·4733	·3367	4·60	·2174	·1537	7·10	·1408	·0995	20·00	·0500	·0354
2·15	·4628	·3289	4·65	·2151	·1521	7·15	·1399	·0989			
2·20	·4527	·3214	4·70	·2128	·1504	7·20	·1389	·0982			
2·25	·4430	·3143	4·75	·2105	·1489	7·25	·1379	·0975			
2·30	0·4337	0·3074	4·80	0·2083	0·1473	7·30	0·1370	0·0969			
2·35	·4247	·3009	4·85	·2062	·1458	7·35	·1361	·0962			
2·40	·4160	·2946	4·90	·2041	·1443	7·40	·1351	·0956			
2·45	·4077	·2886	4·95	·2020	·1428	7·45	·1342	·0949			
2·50	·3996	·2828	5·00	·2000	·1414	7·50	·1333	·0943			

If $f(\xi, \eta)$ is now integrated over the target circle of radius Z, and divided by πZ^2, the expected fraction of the target damaged, E^*, is obtained as

$$E^* = 2\left(\frac{B}{Z}\right)^2\left\{1-\exp\left(-\frac{Z^2}{2(\sigma^2+B^2)}\right)\right\}. \tag{7}$$

To use the approximate formula (7), a method for determining the constant B must be specified. B will be chosen so that the expected damage when placing the diffused damage function at the origin (without random impact) is the same as when placing the cookie-cutter function at the origin. Assume that $R < Z$ both now and throughout the remainder of this paper. Then

$$\pi R^2 = \iint_{\xi^2+\eta^2\leqslant Z^2} \exp\left(-\frac{\xi^2+\eta^2}{2B^2}\right) d\xi\, d\eta$$

and hence

$$R^2 = 2B^2(1-e^{-\frac{1}{2}Z^2/B^2}). \tag{8}$$

Using this result, B/Z can be determined from R/Z or we may write $B = Zf(R/Z)$; numerical results are given in Table 1.

Table 2. *E^* as computed by exponential approximation and by 'cookie-cutter' (Rand Table) for $Z = 1$*

σ	R/Z	E^* from Rand table	E^* from equations (7) and (8)
0·1	0·2	0·0400	0·0400
	·4	·1600	·1596
	·6	·3600	·3560
	·8	·6385	·6330
	·9	·7942	·8135
	·999	·9203	·9984
0·2	0·2	0·0400	0·0400
	·4	·1599	·1577
	·6	·3574	·3436
	·8	·6106	·6120
	·9	·7356	·7976
	·999	·8412	·9982
0·5	0·1	0·0086	0·0087
	·2	·0342	·0337
	·3	·0756	·0735
	·6	·2766	·2632
	·8	·4464	·4947
	·999	·6142	·9969
1·0	0·2	0·0156	0·0155
	·4	·0611	·0594
	·6	·1323	·1334
	·8	·2231	·2899
	·9	·2737	·4896
	·999	·3263	·9923

Because it is intended that this exponential approximation be used in the other two coverage problems described in § 1, it is interesting to determine the adequacy of the approximation. In Table 2, the exponential approximation given in formulae (7) and (8) is compared with the value of E^* obtained from the tables of $S(Z, R)/(\pi Z^2)$ found in Germond (1950).

The data indicate good agreement for $R/Z < 0{\cdot}8$ and for values of σ such that $\sigma/Z \leqslant 1$. Agreement is generally better for smaller σ's and R/Z values, such as will prevail in the cluster and salvo problems.

In the cluster problem (Problem 2) and in the independently aimed salvo (Problem 3), the exponential approximation will be replacing a 'cookie-cutter' damage circle that is aimed off-centre. We may ask how adequately the approximation will fit in this case. The formula (7) is easily extended to give the coverage of an impact circle aimed at an offset point (a, b). In place of (6) we have

$$f(\xi, \eta) = \frac{B^2}{\sigma^2 + B^2} \exp\left(-\frac{(\xi - a)^2 + (\eta - b)^2}{2(\sigma^2 + B^2)}\right). \tag{9}$$

The fraction of coverage, E^*, is then obtained by integrating $f(\xi, \eta)$ over the circle of radius Z and dividing by πZ^2. In fact, using the definition of equation (2) for $p(R, r)$ we have

$$E^* = 2\left(\frac{B}{Z}\right)^2 p(R, r), \tag{10}$$

where $$R^2 = Z^2/(\sigma^2 + B^2), \quad r^2 = (a^2 + b^2)/(\sigma^2 + B^2).$$

As mentioned earlier, $p(R, r)$ is tabled in Rand Corporation (1952), and hence E^* can be computed. However, there appear to be no tabled values for this offset circle *v.* circle problem against which the values of E^* computed by (10) might be compared. In view of this, several hand-run Monte Carlo trials were performed.

Using the Rand set of random normal deviates (Rand Corporation, 1955) and large paper with a grid of lattice points, four off-centre circle *v.* circle coverage experiments were tried. The E^* obtained is compared in Table 3 with E^* as computed from formula (10). The estimation of overlap area for each impact was obtained by counting the number of lattice points covered in the target circle. In addition, four Monte Carlo trials (indicated by p) were conducted in which the overlap area was planimetered.

Table 3. *Comparison of computed E^* with Monte Carlo E^* for eight cases. Target circle of radius $Z = 1$*

Offset distance	R	σ	R/Z	No. of trials	Monte Carlo E^*	E^* computed by (10)
0·33(p)	0·50	0·08	0·50	100	0·250 ± 0·001	0·237
·33(p)	·50	·50	·50	75	·189 ± ·019	·173
·50	·40	·17	·40	75	·152 ± ·008	·144
·67(p)	·33	·06	·33	100	·108 ± ·001	·099
·67	·60	·17	·60	25	·264 ± ·026	·256
·67(p)	·50	·33	·50	50	·170 ± ·021	·163
·83	·20	·17	·20	50	·030 ± ·004	·030
1·17	·25	·33	·25	50	·023 ± ·006	·017

The tolerance indicated about the Monte Carlo estimates of E^* are based upon the sample range (see Burington & May, 1953, p. 167). These tolerances show the uncertainty due to Monte Carlo, but not the error in determining overlap area. The indicated tolerance in Table 3 is $\pm 3s/\sqrt{N}$, where N is the number of Monte Carlo trials and s is the estimated standard deviation.

Assuming an independent elliptical normal distribution of impact points (x, y) about an aim point $(0, 0)$ with standard deviations, σ_x and σ_y in the x and y directions, one can obtain the following for the exponential damage function

$$f(\xi, \eta) = \frac{B^2}{\sqrt{(\sigma_x^2+B^2)}\sqrt{(\sigma_y^2+B^2)}} \exp\left\{-\frac{1}{2}\left(\frac{\xi^2}{\sigma_x^2+B^2}+\frac{\eta^2}{\sigma_y^2+B^2}\right)\right\}.$$

Then this gives

$$E^* = 2\left(\frac{B}{Z}\right)^2\left\{\frac{1}{2\pi\sqrt{(\sigma_x^2+B^2)}\sqrt{(\sigma_y^2+B^2)}}\iint_{\xi^2+\eta^2\leqslant Z^2} \exp\left[-\frac{1}{2}\left(\frac{\xi^2}{\sigma_x^2+B^2}+\frac{\eta^2}{\sigma_y^2+B^2}\right)\right]d\xi\, d\eta\right\}. \quad (11)$$

The expression in braces involves the integration of an elliptical normal function over a co-centred circular region. Approximations for this expression have been found by Oberg (1947); Harter (1959) has provided tables. Fettis (1957, p. 13) has shown that upon setting

$$t = \frac{Z}{\sqrt{(\sigma_x^2+B^2)}}, \quad c = \sqrt{\frac{\sigma_y^2+B^2}{\sigma_x^2+B^2}} \leqslant 1$$

(and switching roles of σ_x, σ_y, if $\sigma_y > \sigma_x$) the probability mass over the circle of radius Z is

$$\left[q\left\{\frac{t}{2}\left(\frac{1}{c}-1\right), \frac{t}{2}\left(\frac{1}{c}+1\right)\right\} - q\left\{\frac{t}{2}\left(\frac{1}{c}+1\right), \frac{t}{2}\left(\frac{1}{c}-1\right)\right\}\right], \quad (12)$$

where $q(R, r)$ is the offset circle probability function previously defined.

3. Models for cluster coverage (Problem 2)

A. *Replacement of cluster with an annulus*

It seems reasonable to replace the cluster of non-overlapping damage circles with an annulus of equivalent area. This automatically includes any rotation of the cluster, i.e. variation in θ (Fig. 4). In the examples below, the mean annulus radius is made to coincide with λ, the radial distance from cluster centre of mass to the centre of any cluster circle.

If there are n circles in the cluster and $2A$ is the annulus width, the equation for equating damage areas and determining A is

$$n\pi R^2 = \pi[(\lambda+A)^2-(\lambda-A)^2] \quad \text{or} \quad A = \frac{nR^2}{4\lambda}. \quad (13)$$

Using the exponential approximation twice, we may approximate the expected coverage by

$$E^* = 2\left[\left(\frac{B_1}{Z}\right)^2\left(1-\exp\left\{-\frac{Z^2}{2(\sigma^2+B_1^2)}\right\}\right) - \left(\frac{B_2}{Z}\right)^2\left(1-\exp\left\{-\frac{Z^2}{2(\sigma^2+B_2^2)}\right\}\right)\right], \quad (14)$$

where

$$B_1 = Zf\left(\frac{\lambda+(nR^2/4\lambda)}{Z}\right), \quad B_2 = Zf\left(\frac{\lambda-(nR^2/4\lambda)}{Z}\right),$$

the function f being defined on p. 315 above.

Table 4 shows the result of some Monte Carlo trials in cases that were run to check the accuracy of formula (14). In each case a cluster of four tangential damage circles of radius $R = \sqrt{0{\cdot}4}$ were employed. This meant that $\lambda = 0{\cdot}4$. Z was chosen as 1. The uncertainty given for the Monte Carlo value is again computed as $\pm 3s/\sqrt{N}$, s being estimated from the sample range. Lattice-point counting was the basis for estimating three areas of coverage and may

have introduced some error. In addition, two Monte Carlo trials (indicated by p) were run with overlap measured by planimeter. Agreement is just about acceptable. However, it appears that the formula is biased in giving estimates that are about 5 to 9 % low. This same phenomenon, already detectable in Table 3, will be seen to occur again in other approximations involving the exponential expression. Perhaps it is characteristic of the spreading out of the damage mass, and could be adjusted for by a constant.

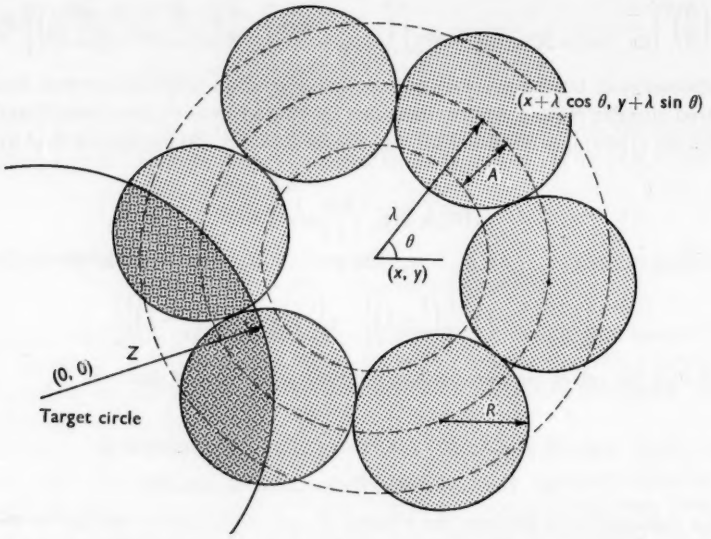

Fig. 4. Replacement of cluster of six circles by an annulus.

Table 4. *Comparison of five Monte Carlo cluster trials with the exponential model* ($n = 4$, $\lambda = 0{\cdot}4$, $R = \sqrt{0{\cdot}4}$, $Z = 1$)

σ	No. of trials	Monte Carlo value of E^*	Value of E^* computed by formula (14)	Value of E^* computed by formula (22)
0·20(p)	50	0·317 ± 0·004	0·302	0·311
·25	30	·308 ± ·012	·294	·303
·40	40	·276 ± ·020	·257	·267
·60(p)	50	·221 ± ·030	·202	·206
·80	30	·163 ± ·042	·153	·154

B. *Rotating cluster model*

If the centre of mass impacts at (x, y), a fixed damage circle will impact at

$$(x+\lambda\cos\theta,\ y+\lambda\sin\theta).$$

Then $P\{(\xi,\eta)$ is damaged|impact of centre mass at $(x,y)\}$

$$= \exp\left(-\frac{(\xi-x-\lambda\cos\theta)^2+(\eta-y-\lambda\sin\theta)^2}{2B^2}\right)$$

and

$$f(\xi,\eta,\lambda,\theta)=\frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\left(-\frac{(\xi-x-\lambda\cos\theta)^2+(\eta-y-\lambda\sin\theta)^2}{2B^2}\right)\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)dx\,dy$$

$$=\frac{B^2}{\sigma^2+B^2}\left\{\exp\left(-\frac{(\xi-\lambda\cos\theta)^2+(\eta-\lambda\sin\theta)^2}{2(\sigma^2+B^2)}\right)\right\}. \tag{15}$$

If θ is uniformly distributed, the *average* damage for the circle is

$$f(\xi,\eta,\lambda)=\frac{B^2}{2\pi(\sigma^2+B^2)}\exp\left(-\frac{\lambda^2+\xi^2+\eta^2}{2(\sigma^2+B^2)}\right)\sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{\lambda}{\sigma^2+B^2}\right)^m\int_0^{2\pi}[\xi\cos\theta+\eta\sin\theta]^m\,d\theta. \tag{16}$$

In the expansion, set $m=2j$ and use the following from Watson (1952, p. 21)

$$\int_0^{2\pi}[\xi\cos\theta+\eta\sin\theta]^m\,d\theta=\begin{cases}0 & \text{if } m \text{ is odd,}\\ \dfrac{2\Gamma(\frac{1}{2}(m+1))\,\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2}(m+2))}(\xi^2+\eta^2)^{\frac{1}{2}m} & m \text{ even.}\end{cases}$$

Then,

$$f(\xi,\eta,\lambda)=\frac{B^2}{\pi(\sigma^2+B^2)}\exp\left(-\frac{\lambda^2+\xi^2+\eta^2}{2(\sigma^2+B^2)}\right)\sum_{j=0}^{\infty}\frac{1}{(2j)!}\left(\frac{\lambda}{\sigma^2+B^2}\right)^{2j}\frac{\Gamma(j+\frac{1}{2})\,\Gamma(\frac{1}{2})}{\Gamma(j+1)}[\xi^2+\eta^2]^j$$

$$=\frac{B^2}{\sigma^2+B^2}\exp\left(-\frac{\lambda^2}{2(\sigma^2+B^2)}\right)\sum_{j=0}^{\infty}\left(\frac{1}{j!}\right)^2\left(\frac{\lambda}{2(\sigma^2+B^2)}\right)^{2j}[\xi^2+\eta^2]^j\exp\left(-\frac{\xi^2+\eta^2}{2(\sigma^2+B^2)}\right). \tag{17}$$

To find the expected damage from this particular damage circle, $f(\xi,\eta,\lambda)$ must be integrated over the target circle of radius Z. To include the effect of n damage circles with damage densities $f_1, f_2, \ldots, f_n$, find an equivalent damage density (assuming independence)

$$g(\xi,\eta,\lambda)=1-\prod_{i=1}^{n}(1-f_i(\xi,\eta,\lambda)) \tag{18}$$

and integrate this over the target circle.

However, the actual physical cluster problem does not allow any overlap of damage circles. The exponential approximation, while it does have overlap, has the impact centres so far apart in the cluster case that the product in (18) is nearly always close to the particular value $(1-f_i)$ of that damage circle nearest to (ξ,η). We shall assume, therefore, that

$$g(\xi,\eta,\lambda)\doteq\sum_{i=1}^{n}f_i(\xi,\eta,\lambda),$$

and hence

$$E^*(\lambda)=\frac{n}{\sigma^2+B^2}\left(\frac{B}{Z}\right)^2\exp\left(-\frac{\lambda^2}{2(\sigma^2+B^2)}\right)\sum_{j=0}^{\infty}\left(\frac{1}{j!}\right)^2\left(\frac{\lambda^2}{2(\sigma^2+B^2)}\right)^j$$

$$\times\left\{\frac{1}{\pi}\iint_{\xi^2+\eta^2\leqslant Z^2}\left[\frac{\xi^2+\eta^2}{2(\sigma^2+B^2)}\right]^j\exp\left(-\frac{\xi^2+\eta^2}{2(\sigma^2+B^2)}\right)d\xi\,d\eta\right\}. \tag{19}$$

The integral in braces can be obtain by a recursion formula. Denoting the term in braces by $g_j(Z)$, we find

$$g_j(Z)=2(\sigma^2+B^2)\int_0^{\frac{1}{2}Z^2/(\sigma^2+B^2)}u^j\,e^{-u}\,du.$$

Setting $h_j(Z)=[2(\sigma^2+B^2)]^{-1}g_j(Z)$ and integrating by parts,

$$h_j(Z)=jh_{j-1}(Z)-\left(\frac{Z^2}{2(\sigma^2+B^2)}\right)^j\exp\left(-\frac{Z^2}{2(\sigma^2+B^2)}\right). \tag{20}$$

In particular

$$h_0(Z) = 1 - \exp\left(-\frac{Z^2}{2(\sigma^2+B^2)}\right). \tag{21}$$

Using the definitions (20) and (21), (19) takes the form

$$E^*(\lambda) = 2n\left(\frac{B}{Z}\right)^2 \exp\left(-\frac{\lambda^2}{2(\sigma^2+B^2)}\right) \sum_{j=0}^{\infty} \left(\frac{1}{j!}\right)^2 \left(\frac{\lambda^2}{2(\sigma^2+B^2)}\right)^j h_j(Z) \leqslant 2n\left(\frac{B}{Z}\right)^2. \tag{22}$$

Although this formula involves an infinite series, it is easy to use in computation. In most cases, after six or so terms, the additional terms may be neglected.

The last column of Table 4 presents the comparison of five estimates produced by this model with the Monte Carlo values. The agreement is better than for the annulus model.

4. Models for salvo coverage with a spread of aim points (Problem 3)

When several damage circles are independently aimed at the target circle, each with its own aim point, the problem of estimating damage is complicated by overlapping (Fig. 3).

Garwood (1947) found the variance of the overlap when n circles are aimed at a target circle, the centres of the circles being uniformly distributed over a larger circle containing the target circle. Bronowski & Neyman (1945) found the mean and the variance of the measure of the area covered by randomly tossing a rectangle n times at a target rectangle. Robbins (1945) has considered the analogous problem for a random sum of n-dimensional integrals in n-space. Walsh (1960) considers selecting an optimal aim point for a salvo of n rounds fired independently at the aim point so as to maximize the kill probability. Germond and Dishington have considered this problem in several Rand Corporation reports.

The damage density function due to the circle aimed at (a_i, b_i) is approximated by

$$f(\xi, \eta, a_i, b_i) = \frac{B^2}{\sigma^2+B^2} \exp\left(-\frac{(\xi-a_i)^2+(\eta-b_i)^2}{2(\sigma^2+B^2)}\right). \tag{23}$$

The damage probability of point (ξ, η) because of effects of the n independent circles is then approximately given by

$$g(\xi, \eta) = 1 - \prod_{i=1}^{n} \left\{1 - \frac{B^2}{\sigma^2+B^2} \exp\left(-\frac{(\xi-a_i)^2+(\eta-b_i)^2}{2(\sigma^2+B^2)}\right)\right\}.$$

Multiplying out and making an obvious partitioning in the exponents

$$g(\xi, \eta) = \sum_{s=1}^{n} (-1)^{s-1} \left(\frac{B^2}{\sigma^2+B^2}\right)^s \left(\sum_{\substack{\text{all } i_1, i_2, \ldots, i_s \\ 1 \leqslant i_1 < i_2 \ldots < i_s}}^{n} \exp\left\{-\frac{\sum_{j=1}^{s} a_{i_j}^2 - \left(\frac{1}{s}\right)\left(\sum_{j=1}^{s} a_{i_j}\right)^2 + \sum_{j=1}^{s} b_{i_j}^2 - \left(\frac{1}{s}\right)\left(\sum_{j=1}^{s} b_{i_j}\right)^2}{2(\sigma^2+B^2)}\right\} \right.$$

$$\left. \times \exp\left\{-\frac{\left(\xi - \sum_{j=1}^{s} a_{i_j}/s\right)^2 + \left(\eta - \sum_{j=1}^{s} b_{i_j}/s\right)^2}{2(\sigma^2+B^2)/s}\right\} \right). \tag{24}$$

The expected damage is now obtained by integrating over the circle of radius of Z. Division by πZ^2 then gives E^*:

$$E^* = \frac{2}{Z^2} \sum_{s=1}^{n} (-1)^{s-1} \left(\frac{B^2}{\sigma^2+B^2}\right)^s \left(\frac{\sigma^2+B^2}{s}\right) \sum_{\substack{\text{all } i_1, i_2, \ldots, i_s \\ 1 \leqslant i_1 < i_2 \ldots < i_s}}^{n}$$

$$\times \exp\left\{-\frac{\sum_{j=1}^{s} a_{i_j}^2 - \left(\frac{1}{s}\right)\left(\sum_{j=1}^{s} a_{i_j}\right)^2 + \sum_{j=1}^{s} b_{i_j}^2 - \left(\frac{1}{s}\right)\left(\sum_{j=1}^{s} b_{i_j}\right)^2}{2(\sigma^2+B^2)}\right\} p(R_s, r_{i_1, i_2, \ldots, i_s}), \tag{25}$$

where

$$R_s^2 = \frac{sZ^2}{\sigma^2+B^2}, \quad r^2_{i_1,i_2,\ldots,i_s} = \frac{\left(\sum_{j=1}^{s} a_{i_j}\right)^2 + \left(\sum_{j=1}^{s} b_{i_j}\right)^2}{s(\sigma^2+B^2)}.$$

The functions $p(R_s, r_{i_1,i_2,\ldots,i_s})$ are the offset circular probability functions defined in equation (2) which may be determined from the Rand Corporation Tables (1952).

Some remarks are necessary concerning the value of B in (25). The expression (8) that related R, Z, and B was basically a demand that the 'cookie-cutter' and the diffused damage function achieve the same expected damage. Because only one damage circle was involved, there was no question of overlap. Also, in the case of the cluster, the individual circles were so far apart that a value for B estimated for each circle as though it were the only circle, seemed to be sufficiently accurate. If the aim points (a_i, b_i) are chosen so that little expected overlap occurs, the value of B given by (8) will again be satisfactory.

Table 5. *Comparison of Monte Carlo salvo trials with the exponential model* (25) ($n = 4$ *circles*, $p = 0{\cdot}80$ *for lattice-point cases*)

R/Z	Z	σ	R	Aim points	No. of Monte Carlo trials	Monte Carlo value of E^*	E^* computed from (25) and (8)
0·258	5	1·5	1·29	(2·5, 0) (0, 2·5) (−2·5, 0) (0, −2·5)	20	0·181 ± 0·017	0·184
·324	5	1·5	1·62	(2·5, 0) (0, 2·5) (−2·5, 0) (0, −2·5)	20	·271 ± ·027	·272
·40(p)	5	1·0	2	(2·5, 0) (0, 2·5) (−2·5, 0) (0, −2·5)	30	·543 ± ·042	·569
·252	10	1·0	2·52	(5·75, 0) (0, 5·75) (−5·75, 0) (0, −5·75)	20	·203 ± ·0004	·201
·30(p)	10	2·0	3	(5, 0) (0, 5) (−5, 0) (0, −5)	30	·327 ± ·013	·339

In the extreme case when $(a_i, b_i) = (0, 0)$ $(i = 1, 2, \ldots, n)$, the expected damage when n 'cookie-cutter' damage functions are centred at the origin is only πR^2. Using diffused exponential damage functions, however, the equation for determining B is

$$\pi R^2 = \iint_{\xi^2+\eta^2 \leqslant Z^2} \left\{1 - \left\{1 - \exp\left(-\frac{\xi^2+\eta^2}{2B^2}\right)\right\}^n\right\} d\xi\, d\eta. \tag{26}$$

If a table similar to Table 1 were made relating R, Z, and B using (26), it would be possible to check the exponential approximation of formula (25) against Dishington's (1956) tables.

Formula (25) with B estimated from (8) was used to compute the expected fraction of coverage in several examples, and the computed values compared with the results of some available Monte Carlo trials. These data appear in Table 5, the intervals of uncertainty being $\pm 3s/\sqrt{N}$, as in the earlier tables. Lattice-point counting was used to estimate areas in three cases, and two cases were planimetered (indicated by (p)).

The lattice-point Monte Carlo data include a nonabort or reliability factor: $p = 0{\cdot}80$. Expression (23) can be multiplied by p to give the damage probability at (ξ, η), also considering reliability. Then (25) would have

$$(-1)^{s-1}\left(\frac{B^2}{\sigma^2+B^2}\right)^s \quad \text{replaced by} \quad (-1)^{s-1}\left(\frac{pB^2}{\sigma^2+B^2}\right)^s.$$

Agreement in Table 5 is probably well within input inaccuracies for any practical applications of the model. Linear interpolation of offset probability function values from tables of Rand Corporation (1952) was used.

REFERENCES

BRONOWSKI, J. & NEYMAN, J. (1945). The variance of the measure of a two-dimensional random set. *Ann. Math. Statist.* **16**, 330–41.

BURINGTON, R. S. & MAY, D. C. (1953). *Handbook of Probability and Statistics with Tables.* Sandusky, Ohio: Handbook Publishers, Inc.

DISHINGTON, R. H. (1956). *The Expected Coverage of a Small Circular Target by a Number of Circular Bombs*, RM-413. The Rand Corporation, Santa Monica, California.†

FETTIS, H. E. (1957). *Some Mathematical Identities and Numerical Methods Relating to the Bivariate Normal Probability for Circular Regions.* Aeronautical Research Laboratory, Wright Air Development Center, *Technical Note*, 57–383.

GARWOOD, F. (1947). The variance of the overlap of geometrical figures with reference to a bombing problem. *Biometrika*, **34**, 1–17.

GERMOND, H. H. (1950). *The Circular Coverage Function*, RM-330. The Rand Corporation, Santa Monica, California.

HARTER, H. L. (1959). *Circular Error Probabilities.* Applied Mathematics Research Branch, Aeronautical Research Laboratory, Wright Air Development Center.

LAURENT, A. G. (1957). Bombing problems. *Opns. Res.* **5**, 75–89.

OBERG, E. N. (1947). Approximate formulas for the radii of circles which include a specified fraction of normal bivariate distribution. *Ann. Math. Statist.* **18** (3), 442–7.

ROBBINS, H. E. (1945). On the measure of a random set. *Ann. Math. Statist.* **16**, 342–7.

RAND CORPORATION (1952). *Offset Circle Probabilities*, R-234. Santa Monica, California.

RAND CORPORATION (1955). *A Million Random Digits with 100,000 Normal Deviates.* Glencoe, Illinois: The Free Press.

THOMPSON, W. A., Jr. (1958). Two uses of the diffused-target concept in aerial target analysis. *Opns. Res.* **6**, 671–5.

WALSH, J. E. (1960). Optimum ammunition properties for salvos. *Opns. Res.* **4**, 204–12.

WATSON, G. N. (1952). *A Treatise on the Theory of Bessel Functions.* Cambridge University Press.

WEIL, H. (1954). The distribution of radial error. *Ann. Math. Statist.* **25**, 168–70.

† Unclassified Rand Corporation reports are on file at several document repositories with leading universities throughout the United States. Generally, they may be obtained through the usual inter-library loan systems.

Biometrika (1961), **48**, 3 *and* 4, *p.* 325
Printed in Great Britain

Ordered tests in the analysis of variance

BY D. J. BARTHOLOMEW
University College of Wales, Aberystwyth

1. INTRODUCTION

In the analysis of variance for a one-way classification it is customary to test the hypothesis that the samples have come from populations with the same mean. The object of the present paper is to discuss the best way of making this test when the unknown population means are subject to order restrictions. A general theory of tests of homogeneity for means under ordered alternatives has been developed in three earlier papers (Bartholomew, 1959*a*, *b*, 1961): these are referred to in this paper as I, II and III, respectively. However, apart from a brief discussion in I, this earlier work relates to the case where the population standard deviations are known *a priori*. This is a natural assumption to make when investigating the theoretical properties of tests for means but is unrealistic in many practical applications. In particular, in the analysis of variance the standard deviations are not usually known, although they can be estimated from the data. In § 2 the $\bar{\chi}^2$-test, first introduced in I and extended in III, will be generalized to cover this case also. The tests based on scores discussed in III may also be adapted for use in the analysis of variance. It will be shown that a close link exists between the latter and the distribution-free test proposed by Jonckheere (1954). A number of factors must be taken into account when choosing a test of which power is one of the most important. Asymptotic results have been obtained for all the tests mentioned above and they are used in § 3 to make power comparisons.

It is instructive to examine more closely the kind of practical situation which gives rise to ordered alternatives in the analysis of variance. The possibility of ranking the class means implies that they correspond to different levels of one or more underlying variates. To take a simple example, suppose that there are k classes with means $\mu_1, \mu_2, \ldots, \mu_k$ with the structure

$$\mu_i = \alpha + \beta x_i \quad (i = 1, 2, \ldots, k),$$

where α and β are unknown constants and $x_1, x_2, \ldots, x_k$ are independent variables for which ranks only are available. The test for equality of means is also a test that β, the regression coefficient, is zero. The ranking of the x's enables us to rank the μ's. Regression models involving more than one independent variable, which can be ranked but not measured, will lead to examples of partial ordering among the μ's. This interpretation is not, of course, confined to *linear* regression models. Suppose that $\mu_i = f(x_i)$ $(i = 1, 2, \ldots, k)$, where $f(x)$ is a monotonic (increasing or decreasing) function of x. This function determines a ranking of the μ's and we may use our theory to test the hypothesis that $f(x)$ is constant. The theory developed here may thus be looked upon as a link between the analysis of variance and regression analysis.

The social sciences provide a fruitful source of applications for ordered tests. Variables such as 'intelligence', 'social class' and 'degree of satisfaction' do not readily lend themselves to precise measurement but can often be ranked without difficulty. The theory presented in this paper thus provides a useful means of testing hypotheses about regression models without the necessity of finding a suitable scale of measurement for the independent

variables. In practice, of course, the testing of hypotheses is only a preliminary step which serves to select the most promising variables for further study. For the later stages of the analysis it is highly desirable that a suitable means of measurement should be devised but it is not necessary for making progress in the early stages of the investigation.

2. THE $\bar{E}^2$-TEST

2·1. *Derivation of the test.* Suppose that we have k groups of observations such that the ith group consists of n_i members drawn from a normal population with mean μ_i and standard deviation σ ($i = 1, 2, \ldots, k$). Let the jth observation in the ith group be denoted by x_{ij} and the mean of the ith group by $\bar{x}_i$. It is assumed that we have prior information which specifies restrictions of the form $\mu_i \geqslant \mu_j$ between some or all of the μ's. If we now attempt to apply the $\bar{\chi}^2$-test, as described in III, to the k means $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ we immediately meet a difficulty. The weights of the observations are n_i/σ^2 and hence $\bar{\chi}^2$ depends on the unknown σ. One way of overcoming this difficulty is to use the likelihood ratio principle to derive a new test treating σ as unknown. This leads to the criterion

$$\bar{E}^2 = \sum_{i=1}^{k} n_i(\hat{\mu}_i - \bar{x})^2 \Big/ \sum_{i=1}^{k}\sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2,*$$

where

$$\bar{x} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij} \Big/ \sum_{i=1}^{k} n_i$$

and $\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_k$ are the maximum likelihood estimates of $\mu_1, \mu_2, \ldots, \mu_k$. The latter can be obtained in the same way as when σ is known because the weights are proportional to $n_1, n_2, \ldots, n_k$ and the μ's depend only on the relative values of the weights. As explained in I, the averaging process by which these weights are obtained is equivalent to the pooling of certain groups resulting in a 'reduced' problem of l groups. If the means of these groups are denoted by $X_1, X_2, \ldots, X_l$, and the corresponding sample numbers of the pooled groups by $N_1, N_2, \ldots, N_l$ then we have

$$\bar{E}^2 = \sum_{i=1}^{l} N_i(X_i - \bar{X})^2 \Big/ \sum_{i=1}^{k}\sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2, \tag{1}$$

where

$$\bar{X} = \sum_{i=1}^{l} N_i X_i \Big/ \sum_{i=1}^{l} N_i = \bar{x}.$$

$\bar{E}^2$ is thus the ratio of the 'between' groups sum of squares to the total sum of squares for the reduced problem. Large values of $\bar{E}^2$ indicate a departure from the null hypothesis.

2·2. *Distribution of $\bar{E}^2$.* The null hypothesis distribution of $\bar{E}^2$ may be readily obtained from results already established. The conditional distribution of $\sum_{i=1}^{l} N_i(X_i - \bar{X})^2/\sigma^2$, for given l, is that of χ^2 with $l-1$ degrees of freedom (by the theorem of § 3·1 in III). Standard theory then gives the conditional distribution of $\bar{E}^2$ as a β-variable with parameters $\frac{1}{2}(l-1)$ and $\frac{1}{2}(N-l)$, $\left(N = \sum_{i=1}^{l} N_i\right)$. The probabilities $p(l, k)$, defined in I, depend only on the relative weights and are therefore independent of σ^2. If $\bar{E}_0^2$ denotes an observed value of $\bar{E}^2$ then its significance may be judged by computing

$$\Pr\{\bar{E}^2 \geqslant \bar{E}_0^2\} = \sum_{l=2}^{k} p(l, k)\, I_{1-y}\{\tfrac{1}{2}(N-l), \tfrac{1}{2}(l-1)\}, \tag{2}$$

* The symbol $\bar{E}^2$ is suggested by the use of E^2 for this ratio in the unrestricted problem; the addition of the bar, as in $\bar{\chi}^2$, indicates that the means are ordered.

where $I_u(p, q)$ is the incomplete β-function and $y = \overline{E}_0^2$. A numerical example of the application of this test was given in I although, as we shall see below, the theory there was different.

In the analysis of variance it is customary to calculate not E^2 but F, which is the ratio of the 'between' groups mean square to the 'within' groups mean square. For any given table, F and E^2 are related as follows

$$F = \nu_2 E^2/\{\nu_1(1-E^2)\},$$

where ν_1 and ν_2 are the degrees of freedom of the numerator and denominator of F, respectively. The choice of F rather than E^2 is purely a matter of convenience because the two tests are identical. This is not so with their ordered counterparts. Suppose that we use a statistic $\overline{F}$ defined as the F-ratio for the reduced problem, then

$$\begin{aligned}\Pr\{\overline{F} \geqslant \overline{F}_0\} &= \sum_{l=2}^{k} p(l,k) \Pr\{F_{l-1,N-l} \geqslant \overline{F}_0\} \\ &= \sum_{l=2}^{k} p(l,k)\, I_{1-z}\{\tfrac{1}{2}(N-l), \tfrac{1}{2}(l-1)\}, \qquad (3)\end{aligned}$$

where

$$z = (l-1)\overline{F}_0/\{(N-l)+(l-1)\overline{F}_0\}. \qquad (4)$$

Both $\overline{F}$ and $\overline{E}^2$ provide a valid test of the null hypothesis but they use different critical regions. This may be seen by comparing z with $\overline{E}_0^2$ expressed in terms of $\overline{F}_0$, thus

$$\overline{E}_0^2 = (l_0-1)\overline{F}_0/\{(N-l_0)+(l_0-1)\overline{F}_0\}, \qquad (5)$$

where l_0 is the *observed* value of l. Equations (4) and (5) differ in that l is a variable of summation in (4) and a constant in (5). The two results were regarded as equivalent in I (eq. (8)) and in II because of a failure to make this distinction.*

2·3. *Comparison with the $\overline{F}$-test.* The $\overline{F}$-test, as described above, may be looked upon as an $\overline{E}^2$-test for which the rejection level depends upon l. In III we showed how the power of ordered tests depended on the choice of critical region when σ was known. The results suggested that the likelihood ratio test, for which the rejection level is constant, could not be substantially improved upon. We may therefore expect that $\overline{E}^2$ will prove to be more powerful than $\overline{F}$. It is, in fact, possible to make large sample power comparisons using existing theory. In the limiting case, as $N \to \infty$, k fixed, the analysis of variance problem treated here and that for known σ discussed previously become equivalent because σ^2 is estimated with infinitely many degrees of freedom. The results given in Table 1, for complete ordering of the μ's, were obtained using this equivalence. Some of these results have already been given in III but have been reproduced here for ease of comparison. The power has been tabulated as a function of

$$\Delta = \sqrt{\left\{\sum_{i=1}^{k} (\mu_i - \overline{\mu})^2\right\}}\Big/\sigma \qquad \left(\overline{\mu} = \frac{1}{k}\sum_{i=1}^{k} \mu_i\right).$$

For reasons given in III the power also depends upon the configuration of the μ's; two extreme cases have therefore been considered. It has been assumed that $n_1 = n_2 = \ldots = n_k$ which makes the weights equal. It will be seen that although $\overline{F}$ is an improvement over F, it is much inferior to $\overline{E}^2$. The difference is so large that we may confidently expect the conclusion to apply when the sample sizes are finite.

* I am indebted to Mr J. E. K. Smith of the Massachusetts Institute of Technology for first drawing my attention to this point.

Table 1. *The asymptotic power functions of* $\bar{E}^2$, $\bar{F}$ *and* F *when* $\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_k$

		Δ				
		0	1	2	3	4
$k=3$	$\bar{E}^2$	0·050	0·244	0·605	0·892	0·987
		·050	·221	·569	·872	·983
	$\bar{F}$	0·050	0·198	0·488	0·799	0·960
		·050	·196	·515	·835	·972
	F	0·050	0·130	0·402	0·776	0·959
$k=4$	$\bar{E}^2$	0·050	0·239	0·594	0·885	0·980
		·050	·202	·531	·849	·978
	$\bar{F}$	0·050	—	—	—	—
		·050	0·168	0·454	0·785	0·957
	F	0·050	0·115	0·350	0·710	0·945

The upper figure of each pair corresponds to equal spacing of the μ's and the lower figure to the case when all but one of the μ's are equal.

3. Other tests

3·1. *Tests based on scores.* We have already pointed out that our problem has close links with regression theory. This correspondence can be exploited if the unknown μ's are replaced by scores satisfying the same order restrictions. This approach has been discussed by Armitage (1955), in an unpublished report by Abelson & Tukey and in III. If σ were known in the present problem our test would be based on

$$T = \sum_{i=1}^{k} c_i n_i \bar{x}_i \Big/ \left(\sum_{i=1}^{k} c_i^2 n_i \right)^{\frac{1}{2}} \sigma, \tag{6}$$

where $c_1, c_2, \ldots, c_k$ are appropriately chosen scores satisfying $\sum_{i=1}^{k} c_i n_i = 0$ and the order restrictions. T is proportional to the regression coefficient of $\bar{x}$ on c; on the null hypothesis it is normally distributed with zero mean and unit variance. The natural modification when σ is unknown is to replace it by the estimate from the within groups mean square of the analysis of variance table. The null hypothesis distribution of T will then be that of t with $\Sigma(n_i - 1)$ degrees of freedom. The power function of the test could be obtained from tables of the non-central t-distribution. The power comparisons with $\bar{\chi}^2$ for equal weights and σ known given in III (Table 9) may be looked upon as large sample results for the analysis of variance problem. It was shown in III that there was little difference between $\bar{\chi}^2$ and the optimum scoring test if k is small and the ranking of the μ's complete. On the other hand, if the order restrictions were not complete, $\bar{\chi}^2$ was much to be preferred. These conclusions apply asymptotically to $\bar{E}^2$ and T. The optimum method of choosing the scores $c_1, c_2, \ldots, c_k$ has been given by Abelson & Tukey for the case of equal weights; some of their results are given in III.

3·2. *Jonckheere's test.* When the μ's can be completely ranked under the alternative hypothesis ($\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_k$) Jonckheere (1954) suggested the following distribution-free test criterion:

$$S = 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (p_{ij} - \tfrac{1}{2} n_i n_j), \tag{7}$$

where
$$p_{ij} = \sum_{r=1}^{n_i} \sum_{s=1}^{n_j} p_{ir,js}$$

and
$$p_{ir,js} = \begin{cases} 1 & \text{if} \quad x_{ir} > x_{js}, \\ 0 & \text{if} \quad x_{ir} \leqslant x_{js} \end{cases} \quad (i < j).$$

Jonckheere showed that S was the appropriate form of Kendall's coefficient of rank correlation when one ranking contains ties. The sampling distribution of S, on the null hypothesis, is approximately normal with zero mean and variance

$$\frac{1}{18}\left\{N^2(2N+3) - \sum_{i=1}^{k} n_i^2(2n_i+3)\right\}.$$

It is instructive to demonstrate the connexion between S and the scoring tests discussed above. Suppose that we put
$$p_{ir,js} = (x_{ir} - x_{js})$$
instead of scoring 0 or 1 according to the sign of the difference. Then we should have the criterion

$$S' = 2\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} n_i n_j \{(\bar{x}_i - \bar{x}_j) - \tfrac{1}{2} n_i n_j\}$$

$$= 2\sum_{i=1}^{k} c_i n_i \bar{x}_i - \text{a constant},$$

where
$$c_i = \sum_{r=i+1}^{k} n_r - \sum_{r=1}^{i-1} n_r \quad (i = 2, 3, \ldots, k-1),$$

$$c_1 = \sum_{r=2}^{k} n_r, \quad c_k = -\sum_{r=1}^{k-1} n_r.$$

Since $c_1 \geqslant c_2 \geqslant \ldots \geqslant c_k$ and $\sum_{i=1}^{k} c_i n_i = 0$, S' is essentially the same as the numerator of (6) and its significance would be judged in the way described for T. If $n_1 = n_2 = \ldots = n_k$, the coefficients c_i are in arithmetic progression and therefore the test is equivalent to the regression coefficient calculated on the assumption that the μ's are equally spaced. Because S takes account only of the sign of the differences between observations we should expect it to be less powerful than S'. On the other hand, it must be remembered that S does not require the assumption of normality.

3·3. *The power of Jonckheere's test.* The power function of S, like that of $\bar{E}^2$, depends upon k parameters but, by specifying the configuration of the μ's, only one need be considered. By assuming that $n_1 = n_2 = \ldots = n_k = n$ here, and throughout this section, we may express the power as a function of

$$\delta = \sqrt{\left\{\sum_{i=1}^{k} (\mu_i - \bar{\mu})^2\right\}} \Big/ \sigma$$

which may be regarded as a measure of the 'distance' of the alternative from the null hypothesis. Let the mean and standard error of S for given δ and n be $\mu_n(S|\delta)$ and $\sigma_n(S|\delta)$, respectively. If δ is fixed and n is allowed to increase without limit, it is clear that the power will tend to 1. However, to obtain comparisons with $\bar{E}^2$ we require the power as a function of

$$\Delta = \sqrt{\left\{\sum_{i=1}^{k} (\mu_i - \bar{\mu})^2\right\}} \Big/ \sqrt{\left(\frac{\sigma^2}{n}\right)} = \delta\sqrt{n},$$

where
$$\sqrt{\left\{\sum_{i=1}^{k} (\mu_i - \bar{\mu})^2\right\}}$$

is divided by the standard error of the means ($\bar{x}_i$) and not by the standard deviation of an individual observation. If Δ is to remain fixed as $n \to \infty$ then δ must tend to zero with $n^{-\frac{1}{2}}$. If we assume that S is asymptotically normally distributed in the neighbourhood of $\delta = 0$, as well as at $\delta = 0$, then it may be shown (Noether, 1955) that as $n \to \infty$

$$\text{Power} \to Q\left\{\nu_\alpha - \Delta \frac{K}{\sigma_n(S|0)}\right\}, \tag{8}$$

where ν_α is the upper $100\alpha\,\%$ point of the unit normal distribution,

$$K = \left.\frac{\partial \mu_n(S|\delta)}{\partial \delta}\right|_{\delta=0} \quad \text{and} \quad Q(x) = \frac{1}{\sqrt{(2\pi)}} \int_x^\infty e^{-\frac{1}{2}u^2}\, du.$$

In order to determine the power we therefore require the expected value of S under the alternative hypothesis. Since the observations are normally distributed with mean μ_i and standard deviation σ,

$$\mathscr{E}(p_{ir,js}) = \Pr\{x_{ir} > x_{js}\} = Q\left\{\frac{\mu_j - \mu_i}{\sigma\sqrt{2}}\right\}.$$

It then follows that

$$\mathscr{E}(p_{ij}) = n^2 Q\left\{\frac{\mu_j - \mu_i}{\sigma\sqrt{2}}\right\}$$

and therefore

$$\mu_n(S|\delta) = 2n^2 \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left\{Q\left(\frac{\mu_j - \mu_i}{\sigma\sqrt{2}}\right) - \frac{1}{2}\right\}. \tag{9}$$

In order to calculate the asymptotic power we must first specify the configuration of μ's in which we are interested and then express (9) in terms of δ. The method will be made clear by considering the calculations for the two cases required for comparison with $\bar{E}^2$.

(*a*) *μ's equally spaced.* Let $\mu_{i-1} - \mu_i = A$ $(i = 2, 3, \ldots, k)$ then $\mu_j - \mu_i = (i-j)A$ $(j > i)$.

$$\sigma^2\delta^2 = \sum_{i=1}^{k} (\mu_i - \bar{\mu})^2 = 2\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (\mu_j - \mu_i)^2 = \frac{k(k^2-1)}{12} A^2,$$

therefore eliminating A we have

$$\mu_j - \mu_i = \delta\sigma(i-j)\sqrt{\left(\frac{12}{k(k^2-1)}\right)}.$$

Substituting in (9)

$$\mu_n(S|\delta) = 2n^2 \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left\{Q\left(\frac{(i-j)\,\delta\sqrt{6}}{\sqrt{[k(k^2-1)]}}\right) - \frac{1}{2}\right\}$$

hence

$$K = n^2\delta\sqrt{\left(\frac{k(k^2-1)}{3\pi}\right)} = n^{\frac{3}{2}}\Delta\sqrt{\left(\frac{k(k^2-1)}{3\pi}\right)}.$$

Now

$$\sigma_n^2(S|0) = n^3 \frac{k(k^2-1)}{9} + O(n^2)$$

so that, in the limit

$$\text{Power} = Q\left(\nu_\alpha - \Delta\sqrt{\frac{3}{\pi}}\right). \tag{10}$$

(*b*) $\mu_1 \geqslant \mu_2 = \ldots = \mu_k$. By a similar argument it may be shown that the asymptotic power in this case is given by

$$\text{Power} = Q\left(\nu_\alpha - \Delta\sqrt{\frac{3}{\pi}}\sqrt{\frac{3}{k+1}}\right). \tag{11}$$

It is obvious that, unless $k = 2$, the test is more powerful in detecting a given Δ when the means are equally spaced than when all but one are equal. Further, in the latter case the power decreases as k increases.

It is interesting to compare (10) and (11) with the corresponding results for the S' test which uses the actual differences between pairs of observations. Here we find the powers to be

$$(a)\quad Q(\nu_\alpha - \Delta) \quad \text{and} \quad (b)\quad Q\left(\nu_\alpha - \Delta\sqrt{\frac{3}{k+1}}\right). \tag{12}$$

There is, therefore, a reduction in power through using S instead of S' but it is only a small one. S has the compensating advantage of being independent of the assumption of normality. An alternative way of comparing S and S' is to compute the asymptotic relative efficiency for the two tests; this is the ratio of sample sizes required to attain equal power

Table 2. *The asymptotic power functions of* $\bar{E}^2$, S *and* F *when* $\mu_1 \geqslant \mu_2 \geqslant \ldots \geqslant \mu_k$

		Δ				
		0	1	2	3	4
$k=3$	$\bar{E}^2$	0·050	0·244	0·605	0·892	0·987
		·050	·221	·569	·872	·983
	S	0·050	0·252	0·622	0·901	0·988
		·050	·212	·519	·814	·959
	F	0·050	0·130	0·402	0·776	0·959
$k=4$	$\bar{E}^2$	0·050	0·239	0·594	0·885	0·980
		·050	·202	·531	·849	·978
	S	0·050	0·252	0·622	0·901	0·988
		·050	·187	·448	·734	·917
	F	0·050	0·115	0·350	0·710	0·945
$k=8$	$\bar{E}^2$	0·050	—	—	—	—
		·050	0·191	0·456	0·800	0·973
	S	0·050	0·252	0·622	0·901	0·988
		·050	·140	·303	·519	·730
	F	0·050	0·090	0·249	0·535	0·853
$k=12$	$\bar{E}^2$	0·050	—	—	—	—
		·050	0·178	0·423	0·766	0·963
	S	0·050	0·252	0·622	0·901	0·988
		·050	·120	·240	·406	·592
	F	0·050	0·080	0·205	0·466	0·776

The upper figure of each pair corresponds to equal spacing to the μ's and the lower figure to the case when all but one of the μ's are equal.

for the same alternative (Noether, 1955). For both configurations which we have considered A.R.E. $= 3/\pi = 0{\cdot}95$ and this may be shown to hold for all configurations. This is in fact a generalization of a well-known result for $k = 2$ (see, for example, Mood, 1954). In this case S is the Wilcoxon or Mann-Whitney statistic and S' is asymptotically the same as the two-sample t-test. For $k = 2$ it is known that S' is the most powerful test but this is not true for $k > 2$. Abelson & Tukey have shown that the power given by (12b) can be increased for $k > 3$ by choosing unequally spaced scores.

Asymptotic relative efficiency cannot be used to compare S (or S') with $\bar{E}^2$ because the limiting distribution of the latter is not even approximately normal unless k is very large. Some numerical results for $\bar{E}^2$, S and F are given in Table 2. Some of the figures for $\bar{E}^2$ and

F have already been given in Table 1, but they are reproduced here for ease of comparison. For $k = 3$ and 4 there is little to choose between $\bar{E}^2$ and S; the latter is slightly more powerful than $\bar{E}^2$ when the μ's are equally spaced and slightly less powerful at the other extreme when all but one of the μ's are equal. In view of the distribution-free character of S it should, perhaps, be preferred to $\bar{E}^2$ under these circumstances. The position is altered for larger k. Although S will maintain its superiority for equal spacing it is very much inferior at the other extreme. In fact, in this case, when $k = 8$ and 12, S is worse than F when $\Delta > 2$ and very much worse than $\bar{E}^2$. The upper values of $\bar{E}^2$ for $k = 8$ and 12 are not known but are unlikely to exceed the upper values for S.

4. CONCLUSIONS

In this paper we have derived the $\bar{E}^2$-test for ordered alternatives in the analysis of variance and reviewed a number of existing tests. In an attempt to select the best test we have made asymptotic power comparisons. Use has been made of the fact that, in the limit, some of the analysis of variance tests are equivalent to tests considered in III. Thus $\bar{E}^2$ and $\bar{\chi}^2$, F and χ^2 have the same limiting power. New results have been obtained for S. The main conclusions on relative power are set out below. It must be borne in mind that they are asymptotic results only and that they involve the assumption of normality.

(*a*) $\bar{E}^2$ is always to be preferred to $\bar{F}$ as defined in I.

(*b*) If the μ's can be completely ranked under the alternative there is little to choose between $\bar{E}^2$ and S if k is small (3 or 4), but if k is large (say > 4) S is liable to be worse than F and therefore $\bar{E}^2$ should be used.

To these we may add a conclusion from III which will be true asymptotically here.

(*c*) If there is partial ordering of the μ's, $\bar{E}^2$ is better than tests using Abelson and Tukey's optimum scores (S is not applicable).

REFERENCES

ARMITAGE, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics*, **11**, 375–86.

BARTHOLOMEW, D. J. (1959*a*). A test of homogeneity for ordered alternatives. *Biometrika*, **46**, 36–48.

BARTHOLOMEW, D. J. (1959*b*). A test of homogeneity for ordered alternatives. II. *Biometrika*, **46**, 328–35.

BARTHOLOMEW, D. J. (1961). A test of homogeneity of means under restricted alternatives. *J. R. Statist. Soc.* B, to appear.

JONCKHEERE, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, **41**, 133–45.

MOOD, A. M. (1954). On the asymptotic efficiency of certain non-parametric two-sample tests. *Ann. Math. Statist.* **25**, 514–22.

NOETHER, G. (1955). On a theorem of Pitman. *Ann. Math. Statist.* **26**, 64–8.

Biometrika (1961), **48**, 3 *and* 4, *p.* 333
Printed in Great Britain

Quota fulfilment using unrestricted random sampling

BY D. H. YOUNG*
University College London

1. INTRODUCTION

The problem to be considered arises in the following way. A sample is required from a population, divided into k strata $w_1, w_2, \ldots, w_k$, such that m_i individuals are obtained from w_i $(i = 1, 2, \ldots, k)$. m_i will be termed the specified quota for stratum w_i. Suppose that either due to cost considerations or sampling restrictions it is necessary to sample randomly from the population without regard to stratification. A sequential scheme may then be used in which individuals are randomly and unrestrictedly drawn from the population until all the specified quotas have been obtained. When sampling finishes, the exact quota will have been drawn from one of the strata, the remaining $k-1$ strata having yielded at least their specified quotas. The number of individuals $n \left(n \geqslant \sum_{i=1}^{k} m_i\right)$ which are sequentially drawn in order to obtain $m_1, m_2, \ldots, m_k$ individuals from $w_1, w_2, \ldots, w_k$, respectively, will be called the least sample number for quota fulfilment. Finally, it is assumed that an effectively infinite population, with a proportion p_i of individuals belonging to w_i, is to be sampled $(\Sigma p_i = 1)$. A general extension of the theory to the finite case would present many additional problems and is beyond the scope of the present paper.

It is of interest to note that for the special case $m_i = 1$, $p_i = k^{-1}$ $(i = 1, 2, \ldots, k)$ the sequential procedure for quota fulfilment is analogous to the classical occupancy problem where balls are randomly and independently thrown into k identical compartments until every compartment is occupied. This case was discussed in a recent paper by Barton & David (1959). The sequential sampling scheme for the binomial case $k = 2$, with $m_1 = m$ say and $m_2 = 0$, has also been considered by several writers (e.g. Haldane (1945) and Finney (1949)) as an alternative to the standard method of estimating the frequency of an attribute from a sample of fixed size.

In this paper some of the basic statistical properties of the random variable n will be considered for more general cases. The results will, for example, be of use in demonstrating the effect of variation in the m_i and p_i, for given k, on:

(*a*) The moments and, in particular, the expectation of the total number of individuals in excess of the sum of specified quotas $(M = \Sigma m_i)$ when sampling finishes. This would be useful for comparing the expected cost of the unrestricted sampling procedure with the cost of sampling restricted to the individual strata, assuming known (different) cost functions for the two procedures.

(*b*) The probability that there will be need for additional sampling, after a specified number of individuals has been drawn from the population.

Results of this kind could be applied to give more flexible principles for test procedures of the type described by Johnson (1957), where an 'optimal' random sample is drawn without regard to stratification and any deficiencies then made up by sampling restricted to each deficient stratum.

* Now at the Unilever Research Laboratory, Sharnbrook, Bedford.

The distribution function and moments of n are not in general capable of concise expression. General formulae will be derived first and then some special cases considered in detail. Several approximations to the upper percentage points of the distribution of n are developed, one being based on the distribution of the extreme positive deviate from the sample mean in a sample of independent unit normal variables. The relative accuracy of the approximations are investigated empirically and a useful upper bound for the maximum error in significance level is given for one of them.

2. Distribution functions of the least sample number for quota fulfilment

The distribution function of n may be obtained as follows. In order that sampling finishes after a total of n unrestricted drawings from the population, the specific quotas from $k-1$ of the strata, and exactly m_i-1 individuals from the remaining stratum, w_i say, must be obtained in the first $n-1$ drawings, the nth drawing yielding a further individual from the deficient stratum w_i. Since the joint distribution of the stratum frequencies in a random sample of size $n-1$, is multinomial of order k with parameters $p_1, p_2, \ldots, p_k$, the probability of n is

$$p(n) = \sum_{i=1}^{k} \binom{n-1}{m_i-1} p_i^{m_i}(1-p_i)^{n-m_i} \Sigma \ldots \Sigma \prod_{j \neq i} \frac{(n-m_i)!}{s_j!} \left(\frac{p_j}{1-p_i}\right)^{s_j}$$

$$= \sum_{i=1}^{k} \frac{(n-1)!}{(m_i-1)!} p_i^{m_i} \Sigma \ldots \Sigma \prod_{j \neq i} \frac{p_j^{s_j}}{s_j!}, \tag{1}$$

where the inner multiple summation is over all integer values of s_j $(j = 1, 2, \ldots, k;\ j \neq i)$ subject to the conditions $\Sigma s_j = n - m_i$ and $s_j \geqslant m_j$ for all j. This form of expression is quite convenient for actual computation of $p(n)$, at least for relatively small values of k and m_i. An alternative form, less concise than (1), but more suitable for studying the moments of n, is

$$p(n) = \sum_{i=1}^{k} \binom{n-1}{m_i-1} p_i^{m_i}(1-p_i)^{n-m_i}$$

$$\times \left\{1 + \sum_{t=1}^{k-2} (-1)^t \sum_{j(i)}^{(t)} \sum_{(s_{ju} < m_{ju})}^{(t)} \frac{(n-m_i)!}{(n-m_i-S_t)!} (1-p_i-P_t)^{n-m_i-S_t} \prod_{u=1}^{t} \frac{p_{ju}^{s_{ju}}}{s_{ju}!}\right\} \tag{2}$$

where

$$S_t = \sum_{u=1}^{t} s_{ju}, \quad P_t = \sum_{u=1}^{t} p_{ju},$$

and where $\sum_{j(i)}^{(t)}$ denotes the summation over all possible integer sets of $(j1, j2, \ldots, jt)$ subject to the conditions

$$\left.\begin{aligned} &1 \leqslant ju \leqslant k \quad (u = 1, 2, \ldots, t), \\ &ju \neq i \quad \text{for all } u, \\ &ju \neq ju' \quad \text{if} \quad u \neq u', \end{aligned}\right\} \tag{3}$$

and $\sum_{(s_{ju} < m_{ju})}^{(t)}$ represents the multiple sum over integer values $s_{j1}, s_{j2}, \ldots, s_{jt}$ subject to $s_{ju} \leqslant m_{ju} - 1$ $(u = 1, 2, \ldots, t)$. Equation (2) is valid for integer values of $k \geqslant 2$, provided $\sum_{t=1}^{0}$ is interpreted as zero. For the symmetric case $m_i = m$, $p_i = k^{-1}$ $(i = 1, 2, \ldots, k)$, (2) simplifies to

$$p(n) = \binom{n-1}{m-1} \frac{(k-1)^{n-m}}{k^{n-1}} + \frac{(n-1)!}{(m-1)!} k^{-(n-1)} \sum_{t=1}^{k-2} (-1)^t \binom{k-1}{t} \sum_{(s_u < m)}^{(t)} \frac{(k-t-1)^{n-m-S_t}}{s_1! \ldots s_t! (n-m-S_t)!}. \tag{4}$$

If in (4) we put $m = 1$ we obtain the well-known results for the classical sequential occupancy problem

$$p(n) = \left(1-\frac{1}{k}\right)^{n-1}\left\{1-\binom{k-1}{1}\left(1-\frac{1}{k-1}\right)^{n-1}+\binom{k-1}{2}\left(1-\frac{2}{k-1}\right)^{n-1}-\ldots\right\} \tag{5}$$

or, in terms of differences of zero,

$$p(n) = \frac{\Delta^{k-1}0^{n-1}}{k^{n-1}}. \tag{6}$$

The probability distribution of n also simplifies a great deal for the case $m = 2$ and may be written

$$p(n) = k^{-(n-1)}\sum_{t=0}^{k-2}(-1)^t\binom{k-1}{t}\sum_{i=0}^{t}\binom{t}{i}(n-1)^{i+1}(k-t-1)^{n-2-i}. \tag{7}$$

3. Moments of n

The ascending factorial moments of n may be derived directly from (2) but, as we would expect, they are not capable of concise expression for general m_i and p_i $(i = 1, 2, \ldots, k)$. We will derive the general formula and then proceed to discuss special cases.

From (2), the rth ascending factorial moment of n is given by

$$\begin{aligned}\mathscr{E}(n(n+1)\ldots(n+r-1)) &= \sum_{i=1}^{k}\frac{p_i^{m_i}}{(m_i-1)!}\sum_{n=M}^{\infty}\frac{(n+r-1)!}{(n-m_i)!}(1-p_i)^{n-m_i}\\ &\quad+\sum_{i=1}^{k}\frac{p_i^{m_i}}{(m_i-1)!}\sum_{t=1}^{k-2}(-1)^t\sum_{j(i)}^{(t)}\sum_{\{s_{ju}<m_{ju}\}}^{(t)}\prod_{u=1}^{t}\frac{p_{ju}^{s_{ju}}}{s_{ju}!}\sum_{n=M}^{\infty}\frac{(n+r-1)!}{(n-m_i-S_t)!}(1-p_i-P_t)^{n-m_i-S_t}.\end{aligned}$$

Applying the result for the summation with respect to the index of a binomial term

$$\sum_{N=K}^{\infty}\binom{N}{\omega}p^{\omega}(1-p)^{N-\omega} = p^{-1}I_{1-p}(K-\omega,\omega+1),$$

where in the usual incomplete beta function notation

$$I_p(a, N-a+1) = 1-\sum_{\omega=0}^{a-1}\binom{N}{\omega}p^{\omega}(1-p)^{N-\omega},$$

gives

$$\begin{aligned}\mathscr{E}(n(n+1)\ldots(n+r-1)) &= \sum_{i=1}^{k}\frac{(m_i+r-1)!}{(m_i-1)!}p_i^{-r}I_{1-p_i}(M-m_i, m_i+r)\\ &\quad+\sum_{i=1}^{k}\frac{p_i^{m_i}}{(m_i-1)!}\sum_{t=1}^{k-2}(-1)^t\sum_{j(i)}^{(t)}\sum_{\{s_{ju}<m_{ju}\}}^{(t)}\frac{(m_i+S_i+r-1)!}{(p_i+P_t)^{m_i+r+S_t}}\\ &\quad\times I_{1-p_i-P_t}(M-m_i-S_t, m_i+r+S_t)\prod_{u=1}^{t}\frac{p_{ju}^{s_{ju}}}{s_{ju}!}.\end{aligned} \tag{8}$$

Some specific cases will now be considered for which (8) reduces to fairly concise forms.

(a) $k = 2$, *general* m_i *and* p_i $(i = 1, 2)$

For the binomial case, the rth ascending factorial moment of n is

$$\mathscr{E}(n(n+1)\ldots(n+r-1)) = (m_1+r-1)^{(r)}p_1^{-r}I_{p_2}(m_2, m_1+r)+(m_2+r-1)^{(r)}p_2^{-r}I_{p_1}(m_1, m_2+r). \tag{9}$$

For large m_1 and m_2 standard normal approximations may be applied to the incomplete beta functions, giving

$$\mathscr{E}(n(n+1)\dots(n+r-1)) \doteqdot (m_1+r-1)^{(r)}p_1^{-r}\Phi\left\{\frac{m_1p_2-m_2p_1+rp_2+p_1-\frac{1}{2}}{(m_1+m_2+r-1)^{\frac{1}{2}}(p_1p_2)^{\frac{1}{2}}}\right\}$$
$$+(m_2+r-1)^{(r)}p_2^{-r}\Phi\left\{\frac{m_2p_1-m_1p_2+rp_1+p_2-\frac{1}{2}}{(m_1+m_2+r-1)^{\frac{1}{2}}(p_1p_2)^{\frac{1}{2}}}\right\},$$

where $$\Phi\{x\} = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\,dt.$$

If accurate estimates or even the exact values of the stratum frequencies p_1, p_2 are available before sampling starts, the specified quotas m_1, m_2 may be chosen proportional to p_1, p_2, respectively. In this case

$$\mathscr{E}(n(n+1)\dots(n+r-1)) \doteqdot (m_1+r-1)^{(r)}\left(\frac{m_1+m_2}{m_1}\right)^r\Phi\left\{\frac{m_2(2r-1)+m_1}{2(m_1m_2)^{\frac{1}{2}}(m_1+m_2+r-1)^{\frac{1}{2}}}\right\}$$
$$+(m_2+r-1)^{(r)}\left(\frac{m_1+m_2}{m_2}\right)^r\Phi\left\{\frac{m_1(2r-1)+m_2}{2(m_1m_2)^{\frac{1}{2}}(m_1+m_2+r-1)^{\frac{1}{2}}}\right\}. \quad (10)$$

If the Φ functions are expanded in exponential series and terms of order $m_1^{-\frac{1}{2}}$ or $m_2^{-\frac{1}{2}}$ neglected, further approximations to (10) can be obtained. In particular the mean value of n is approximately

$$\mathscr{E}(n) \doteqdot m_1+m_2+\left(\frac{m_1+m_2}{2\pi m_1m_2}\right)^{\frac{1}{2}} \quad (11)$$

and the variance of n for the symmetric case $m_1 = m_2 = m$ and $p_1 = p_2 = \frac{1}{2}$, is

$$\text{var}\,(n) \doteqdot 2m(1-2\pi^{-1})+2m^{\frac{1}{2}}\pi^{-\frac{1}{2}}+\pi^{-1}. \quad (12)$$

Exact values of the expectation and variance of n have been calculated for the symmetric case, for some typical values of m, and are shown in columns (1) and (3) respectively in Table 1. Columns (2) and (4) give the corresponding approximate values obtained from (11), putting $m_1 = m_2 = m$, and (12), respectively. Table 2 gives values of $\mathscr{E}(n)$ for the case when m_1, m_2 are proportional to p_1, p_2 respectively for $p_1 = 0{\cdot}5\,(0{\cdot}1)\,0{\cdot}9$ and $M = m_1+m_2 = 20, 50, 100$. The figures shown in column (1) are the exact values obtained from (9) putting $r = 1$, while column (2) gives approximate values obtained from (11).

Table 1. *Values of* $\mathscr{E}(n)$ *and* var (n) *for the symmetric case* $p_1 = p_2 = \frac{1}{2}$, $m_1 = m_2 = m$

(1) $\mathscr{E}(n)$: exact; (2) $\mathscr{E}(n)$: approximation (11); (3) var (n): exact; (4) var (n): approximation (12).

m	(1)	(2)	(3)	(4)
5	12·461	12·523	6·405	6·475
10	23·524	23·568	11·106	11·154
15	34·334	34·370	15·551	15·590
20	45·015	45·046	19·866	19·900
30	66·155	66·180	28·275	28·302
40	87·114	87·137	36·502	36·525
50	107·959	107·979	44·615	44·634

The approximations in Tables 1 and 2 tend to overestimate the true values of $\mathscr{E}(n)$ and var (n), but the errors become practically negligible with increasing m_1 and m_2.

Table 2. *Values of $\mathscr{E}(n)$ when m_1, m_2 are proportional to p_1, p_2, respectively*

(1) $\mathscr{E}(n)$: exact; (2) $\mathscr{E}(n)$: approximation (11).

		$m_1+m_2=20$		$m_1+m_2=50$		$m_1+m_2=100$	
p_1	p_2	(1)	(2)	(1)	(2)	(1)	(2)
0·5	0·5	23·524	23·568	55·614	55·642	107·640	107·979
0·6	0·4	23·594	23·642	55·728	55·758	108·132	108·143
0·7	0·3	23·833	23·893	56·117	56·256	108·688	108·706
0·8	0·2	24·364	24·460	56·991	57·052	109·948	109·974
0·9	0·1	25·703	25·947	59·246	59·403	113·239	113·298

(*b*) $k=3$, $m_i=m$ *and* $p_i=\frac{1}{3}$ $(i=1,2,3)$

For the trinomial symmetric case, the rth ascending factorial moment of n may be written

$$\mathscr{E}(n(n+1)\dots(n+r-1)) = 3^{r+1}\Big\{(m+r-1)^{(r)}\, I_{\frac{1}{3}}(2m, m+r) - \sum_{s=0}^{m-1} \frac{(m+s+r-1)^{(r+s)}}{s!\, 2^{m+s+r-1}}\, I_{\frac{1}{3}}(2m-s, m+s+r)\Big\}. \quad (13)$$

In particular for $r=1$, the mean value of n after some algebraic manipulation is

$$\mathscr{E}(n) = 9m \sum_{s=0}^{m} \binom{m+s-1}{m-1} \left(\frac{1}{2}\right)^{m+s-1} I_{\frac{1}{3}}(m, m+s). \quad (14)$$

Table 3. *Values of $\mathscr{E}(n)$ for the symmetric trinomial case*

m	$\mathscr{E}(n)$ Exact (15)	$\mathscr{E}(n)$ Approx. (16)
1	5·50	5·95
2	9·64	10·00
3	13·49	13·81
4	17·19	17·49
5	20·81	21·09
10	38·20	38·44
20	71·56	71·77

An alternative expression, derived by Young (1960), which contains a fixed (small) number of terms, for all m, and is therefore more useful for computational purposes is

$$\mathscr{E}(n) = 3m + 9m\binom{2m}{m} 2^{-2m}\, I_{\frac{1}{3}}(m+2, 2m+1) + (9m+5)\frac{(3m)!\, 3^{-3m}}{(m-1)!\, m!\, (m+1)!}. \quad (15)$$

The following approximation to (15), which neglects terms of order less than or equal to $m^{-\frac{1}{2}}$, is obtained by simple application of Stirling's formula to the factorials and the standard Normal approximation to the incomplete beta function

$$\mathscr{E}(n) \doteqdot 3m + \frac{9}{2}\sqrt{\frac{m}{\pi}} + \frac{3\sqrt{3}}{4\pi}. \quad (16)$$

Table 3, which gives values of $\mathscr{E}(n)$, for some small values of m, indicates the increasing accuracy of (16) with increasing m.

(c) $m_i = 2$, $p_i = k$ $(i = 1, 2, \ldots, k)$

In this case (8) simplifies to

$$\mathscr{E}(n(n+1)\ldots(n+r-1)) = k^{r+1}\sum_{t=0}^{k-2}(-1)^t\binom{k-1}{t}\sum_{j=0}^{t}\binom{t}{j}\frac{(r+j+1)!}{(t+1)^{r+j+2}}I_{(k-t-1)/k}(2k-2-j, r+j+2). \tag{17}$$

Values $\mathscr{E}(n)$ have been calculated from (17), putting $r = 1$, for $k = 2(1)5$ and are shown in column (1) in Table 4. In column (2), corresponding values of $\mathscr{E}(n)$ for the case $m_i = 1$, $p_i = k^{-1}$ $(i = 1, 2, \ldots, k)$ are given calculated from the well-known formula

$$\mathscr{E}(n) = k\sum_{j=1}^{k} j^{-1}. \tag{18}$$

The figures in column (3) are the ratios of expected total 'excess' of individuals for $m = 1$ to the expected total 'excess' for $m = 2$. These ratios appear to be fairly stable, with respect to variation in k.

Table 4

(1) $\mathscr{E}(n)$ when $m = 2$, $p_i = k^{-1}$; (2) $\mathscr{E}(n)$ when $m_i = 1$, $p_i = k^{-1}$ $(i = 1, 2, \ldots, k)$;
(3) $(\mathscr{E}(n|m=1)-k)/(\mathscr{E}(n|m=2)-2k)$.

k	(1)	(2)	(3)
2	5·500	3·000	0·667
3	9·639	5·500	0·687
4	14·186	8·333	0·701
5	19·041	11·417	0·709

4. Upper significance limits for n

Finally, to complete this paper, the problem of determining the upper percentage points of the distribution of n will be considered. An approximation, which is sufficiently accurate to determine the correct integer-valued percentage points in nearly all cases, whatever the values of m_i, p_i $(i = 1, 2, \ldots, k)$ and k, is obtained as follows. Consider the stage in the sequential sampling procedure when n^* individuals $(n^* \geqslant M = \Sigma m_i)$ have been randomly drawn from the population. If n_i of the n^* individuals belong to w_i, then

$$\Pr\{n > n^*\} = \Pr\left\{\sum_{i=1}^{k} E_i\right\}, \tag{19}$$

where E_i denotes the event $n_i < m_i$, i.e.

$$\Pr\{n \geqslant n^*+1\} = \sum_{i=1}^{k}\Pr\{E_i\} - \sum\sum_{i<j}\Pr\{E_iE_j\} + \sum\sum\sum_{i<j<l}\Pr\{E_iE_jE_l\} - \ldots. \tag{20}$$

If this series is terminated, the remainder is less in absolute magnitude than the last term omitted and of the same sign. For the present purpose, it is convenient to use the inequality

$$\sum_{i=1}^{k}\Pr\{E_i\} - \Pr\{n \geqslant n^*+1\} < \sum\sum_{i<j}\Pr\{E_iE_j\},$$

or

$$\sum_{i=1}^{k}\sum_{t_i=0}^{m_i-1}\binom{n^*}{t_i}p_i^{t_i}(1-p_i)^{n^*-t_i} - \Pr\{n \geqslant n^*+1\} < \sum\sum_{i<j}\sum_{t_i=0}^{m_i-1}\sum_{t_j=0}^{m_j-1}\frac{n^*!\,p_i^{t_i}p_j^{t_j}(1-p_i-p_j)^{n^*-t_i-t_j}}{t_i!\,t_j!\,(n^*-t_i-t_j)!}. \tag{21}$$

(21) may be replaced by the weaker inequality

$$\sum_{i=1}^{k}\sum_{t_i=0}^{m_i-1}\binom{n^*}{t_i}p_i^{t_i}(1-p_i)^{n^*-t_i}-\Pr\{n\geqslant n^*+1\}<\sum\sum_{i<j}\sum_{t_i=0}^{m_i-1}\binom{n^*}{t_i}p_i^{t_i}(1-p_i)^{n^*-t_i}\sum_{t_j=0}^{m_j-1}\binom{n^*}{t_j}p_j^{t_j}(1-p_j)^{n^*-t_j}.$$

Hence $$\sum_{i=1}^{k}\sum_{t_i=0}^{m_i-1}\binom{n^*}{t_i}p_i^{t_i}(1-p_i)^{n^*-t_i}-\Pr\{n\geqslant n^*+1\}<\frac{1}{2}\left\{\sum_{i=1}^{k}\sum_{t_i=0}^{m_i-1}\binom{n^*}{t_i}p_i^{t_i}(1-p_i)^{n^*-t_i}\right\}^2. \quad (22)$$

For n^*, sufficiently large, the upper bound on the right-hand side of (22) will be negligible and

$$\Pr\{n\geqslant n^*+1\}\doteqdot\sum_{i=1}^{k}\sum_{t_i=0}^{m_i-1}\binom{n^*}{t_i}p_i^{t_i}(1-p_i)^{n^*-t_i}=\sum_{i=1}^{k}I_{1-p_i}(n^*-m_i+1,m_i). \quad (23)$$

Defining the nominal upper $100\alpha\,\%$ point n_α, of the distribution of n, as the least integer values satisfying the inequality

$$\Pr\{n\geqslant n_\alpha\}\leqslant\alpha$$

and applying (23), approximate solutions for the n_α are obtained as the least integer values for which

$$\sum_{i=1}^{k}I_{1-p_i}(n_\alpha-m_i,m_i)\leqslant\alpha. \quad (24)$$

From (22) it follows that the approximate significance level, as computed by the left-hand side of (24), differs from the exact significance level by less than

$$\frac{1}{2}\left\{\sum_{i=1}^{k}I_{1-p_i}(n_\alpha-m_i,m_i)\right\}^2$$

which, from (24), cannot exceed $\frac{1}{2}\alpha^2$. Clearly for α of the order 0·05, 0·01, this error in significance level will be negligible for most practical purposes.

Developing the argument further, (22) may be applied in the form

$$\Pr\{n\geqslant n_\alpha+1\}>\sum_{i=1}^{k}I_{1-p_i}(n_\alpha-m_i+1,m_i)-\frac{1}{2}\left\{\sum_{i=1}^{k}I_{1-p_i}(n_\alpha-m_i+1,m_i)\right\}^2.$$

The least integer values n_α, as given by (24), must then be the correct *nominal* percentage points if

$$\sum_{i=1}^{k}I_{1-p_i}(n_\alpha-m_i+1,m_i)-\frac{1}{2}\left\{\sum_{i=1}^{k}I_{1-p_i}(n_\alpha-m_i+1,m_i)\right\}^2>\alpha$$

or, neglecting terms of order less than or equal to α^3, if

$$\sum_{i=1}^{k}I_{1-p_i}(n_\alpha-m_i+1,m_i)>\alpha+\tfrac{1}{2}\alpha^2. \quad (25)$$

For m_i small, values of n_α may be found directly from (24), using tables of the incomplete beta function (ed. K. Pearson (1934)). If, however, the m_i are large and the p_i's not too small the incomplete beta functions may be approximated by Normal probability integrals: the n_α are then the least integer values satisfying

$$\sum_{i=1}^{k}\Phi\left\{\frac{m_i-\frac{1}{2}-(n_\alpha-1)p_i}{((n_\alpha-1)p_i(1-p_i))^{\frac{1}{2}}}\right\}\leqslant\alpha. \quad (26)$$

If the inequality in (26) is replaced by an equality and terms of order $m^{-\frac{1}{2}}$ neglected then, for the symmetric case,

$$n_\alpha\doteqdot km+\{k(k-1)n\}^{\frac{1}{2}}\lambda_{\alpha/k}+\tfrac{1}{2}k(\lambda_{\alpha/k}^2-1)-(\tfrac{1}{2}\lambda_{\alpha/k}^2-1), \quad (27)$$

where $\lambda_{\alpha/k}$ is the upper $100\alpha/k$ percentage point of the standard Normal distribution.

For the symmetric case, an alternative approximation to (27) may be used. From (19)

$$\Pr\{n \leqslant n^*\} = \Pr\{n_1 \geqslant m_1, n_2 \geqslant m_2, \ldots, n_k \geqslant m_k\}, \tag{28}$$

n_i being the number of individuals drawn from w_i in a sample of n^* unrestricted drawings from the population.

If $m_1 = m_2 = \ldots = m_k$ and $p_1 = p_2 = \ldots = p_k$ then

$$\Pr\{n \leqslant n^*\} = \Pr\{\min_i n_i \geqslant m\}. \tag{29}$$

Johnson & Young (1960) showed that the joint distribution of the standardized deviates, $(kn_i - N)(N(k-1))^{-\frac{1}{2}}$ of the symmetrical multinomial distribution, with index N and equal parameters k^{-1}, may be approximated by that of the deviates from the sample mean

$$\left(\frac{k}{k-1}\right)^{\frac{1}{2}}(x_1 - \bar{x}), \quad \left(\frac{k}{k-1}\right)^{\frac{1}{2}}(x_2 - \bar{x}), \quad \ldots, \quad \left(\frac{k}{k-1}\right)^{\frac{1}{2}}(x_k - \bar{x}),$$

where $x_1, x_2, \ldots, x_k$ are k independent unit normal variables.

In particular the minimum multinomial frequency is approximately distributed as $Nk^{-1}(1 - k^{\frac{1}{2}}N^{-\frac{1}{2}}u_k)$, where u_k is the extreme positive deviate from the sample mean in a sample of k independent unit normal variables. Applying this in (29) gives

$$\Pr\{n \leqslant n^*\} \doteqdot \Pr\left\{u_k \leqslant \frac{n^* - km + \frac{1}{2}k}{(kn^*)^{\frac{1}{2}}}\right\}. \tag{30}$$

The factor $\frac{1}{2}k$ in (30) arises from the use of a continuity correction $\frac{1}{2}$ which is simply half the interval between the consecutive values of the distribution of n. From (30), it follows that the $100\alpha\,\%$ point n_α of the distribution of n is approximately given by

$$\frac{n_\alpha - 1 - km + \frac{1}{2}k}{(k(n_\alpha - 1))^{\frac{1}{2}}} \doteqdot u_{\alpha,k}, \tag{31}$$

where $u_{\alpha,k}$ is the upper $100\alpha\,\%$ point of the distribution of u_k.

Table 25 of Pearson & Hartley (1958) gives values of $u_{\alpha,k}$ obtained from the probability integral tabulated by Nair (1948). Finally, if terms of order less than or equal to $m^{-\frac{1}{2}}$ are neglected then the solution to (31) is

$$n_\alpha \doteqdot km - \tfrac{1}{2}(k-2) + \tfrac{1}{2}ku_{\alpha,k}^2 + ku_{\alpha,k}\sqrt{m}. \tag{32}$$

In a preliminary investigation, the least integer values n_α satisfying (24) have been calculated for the symmetric case for $k = 2, 3, 4$, $\alpha = 0{\cdot}05$, $0{\cdot}01$ and $m = 1\,(1)\,6$. These are shown in columns (a) of Table 5. The figures given in brackets are the corresponding approximate significance levels as computed by the left-hand side of (24). These differ from the exact significance levels by less than $\frac{1}{2}\alpha^2$. Columns (b) and (c) show approximate values of n_α as calculated from (27) and (32), respectively. These values, although given formally to two decimal places, would, in fact, give practical procedures identical with those given by the next larger integer values.

For $k = 2$, the values of n_α as given in columns (a), (b) and (c) would, in all cases, give identical practical procedures. For $k > 2$, approximations (27) and (32) give too large a value of n_α and hence too small a significance level. However, there are clear indications, especially for $k = 3$, that this is due to the relatively small values of m that are considered in Table 5. Table 6, which has been constructed using (27), gives approximate values of n_α

Table 5. *Upper percentage points of the distribution of n when* $m_i = m$, $p_i = k^{-1}$ $(i = 1, 2, \ldots, k)$

(*a*) Inequality (24), (*b*) approximation (27), (*c*) approximation (32).

		$\alpha = 0{\cdot}05$			$\alpha = 0{\cdot}01$		
	k	(*a*)	(*b*)	(*c*)	(*a*)	(*b*)	(*c*)
$m = 1$	2	7 (0·0312)	6·69	6·69	9 (0·0078)	8·96	8·96
	3	12 (0·0347)	12·24	12·26	16 (0·0069)	16·51	16·55
	4	17 (0·0401)	18·30	18·29	22 (0·0095)	24·54	24·53
$m = 2$	2	10 (0·0391)	9·84	9·84	13 (0·0063)	12·47	12·47
	3	17 (0·0411)	17·40	17·42	21 (0·0099)	22·26	22·31
	4	24 (0·0464)	25·52	25·50	31 (0·0079)	32·57	32·56
$m = 3$	2	13 (0·0386)	12·72	12·72	16 (0·0074)	15·63	15·63
	3	22 (0·0385)	22·06	22·08	27 (0·0075)	27·37	27·43
	4	31 (0·0424)	31·98	31·97	38 (0·0083)	39·66	39·65
$m = 4$	2	16 (0·0352)	15·47	15·47	19 (0·0075)	18·60	18·60
	3	26 (0·0451)	26·45	26·48	32 (0·0073)	32·15	32·21
	4	37 (0·0420)	38·06	38·05	44 (0·0097)	46·27	46·25
$m = 5$	2	18 (0·0490)	17·49	17·49	22 (0·0072)	21·46	21·45
	3	31 (0·0486)	30·68	30·71	36 (0·0090)	36·72	36·78
	4	43 (0·0434)	43·89	43·88	51 (0·0084)	52·56	52·54
$m = 6$	2	21 (0·0414)	20·71	20·71	25 (0·0066)	24·24	24·24
	3	35 (0·0391)	34·80	34·82	41 (0·0077)	41·14	41·20
	4	48 (0·0497)	49·55	49·53			

Table 6. *Approximate upper significance limits for n when* $m_i = m$, $p_i = k^{-1}$ $(i = 1, 2, \ldots, k)$

	$m = 10$		$m = 15$		$m = 20$	
k	0·05	0·01	0·05	0·01	0·05	0·01
2	30·69	34·84	42·66	47·43	54·32	59·61
3	50·51	57·88	69·22	77·60	87·34	96·58
4	71·09	81·57	96·61	108·48	121·26	134·30
5	92·22	105·77	124·62	139·92	155·85	172·63
6	113·79	130·88	153·11	171·81	190·97	211·44
	$m = 25$		$m = 30$		$m = 35$	
k	0·05	0·01	0·05	0·01	0·05	0·01
2	65·78	71·53	77·10	83·27	88·32	98·87
3	105·09	115·09	122·58	133·26	139·87	151·18
4	145·06	159·44	169·06	184·08	192·47	208·35
5	186·34	204·43	216·31	235·57	245·87	266·22
6	227·89	249·93	264·15	287·60	299·90	324·65
	$m = 40$		$m = 45$		$m = 50$	
k	0·05	0·01	0·05	0·01	0·05	0·01
2	99·45	106·36	110·52	117·75	121·52	129·08
3	157·00	168·89	173·99	186·43	190·89	203·85
4	215·64	232·32	238·62	256·05	261·44	279·58
5	275·12	296·48	304·11	326·41	332·89	356·09
6	335·26	361·22	370·29	397·39	405·05	433·22

for $m = 10(5)50$, $k = 2(1)6$ and $\alpha = 0{\cdot}05$, $0{\cdot}01$. For $k = 2$, 3, (27) almost certainly determines the correct integer-valued *nominal* percentage points. For the larger values of k, and m small, the significance limits may be in error by one or two at the most.

REFERENCES

BARTON, D. E. & DAVID, F. N. (1959). Sequential occupancy. *Biometrika*, **46**, 218–23.

FINNEY, D. J. (1949). On a method of estimating frequencies. *Biometrika*, **36**, 233–4.

HALDANE, J. B. S. (1945). On a method of estimating frequencies. *Biometrika*, **33**, 222–25.

JOHNSON, N. L. (1957). Optimal sampling for quota fulfilment. *Biometrika*, **44**, 518–20.

JOHNSON, N. L. & YOUNG, D. H. (1960). Some applications of two approximations to the multinomial distribution. *Biometrika*, **47**, 463–69.

NAIR, K. R. (1948). The distribution of the extreme deviate from the sample mean and its studentized form. *Biometrika*, **35**, 118–44.

PEARSON, K. (1934). *Tables of the Incomplete Beta-Function*. Cambridge University Press.

PEARSON, E. S. & HARTLEY, H. O. (1958). *Biometrika Tables for Statisticians*, **1**, 2nd ed. Cambridge University Press.

YOUNG, D. H. (1960). Unpublished Ph.D. Thesis. University of London.

Biometrika (1961), **48**, 3 *and* 4, *p.* 343
With 1 *text-figure*
Printed in Great Britain

Large-sample estimation of parameters for moving-average models

BY A. M. WALKER
Statistical Laboratory, University of Cambridge

1. INTRODUCTION

Let $\{x_t\}$, $t = 0, \pm 1, \pm 2, \ldots$, be a stationary normal moving-average process, defined by

$$x_t = \mu + \epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_h \epsilon_{t-h}, \tag{1}$$

where (ϵ_t) is a set of independent random variables, each distributed normally with mean zero and variance σ^2. The problem of making inferences about the parameters $\beta_1, \beta_2, \ldots, \beta_h$, given a sample of consecutive observations $(x_1, x_2, \ldots, x_n)$ from the process, is a well-known one in time-series analysis. Little progress with this seems to have been made except under the assumption that n is large, but in the large-sample case the work of Whittle (1951, Chapter 7, 1953, 1954, pp. 211–18) enables one to obtain, at least in principle, a solution which for most purposes can be regarded as complete. From this work it follows that, provided that the roots of the equation $z^h + \beta_1 z^{h-1} + \ldots + \beta_h = 0$ all have moduli less than unity, the logarithm of the likelihood of the sample is given asymptotically by

$$L \sim -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{4\pi}\int_{-\pi}^{\pi} \left| \sum_{t=1}^{n} (x_t - \mu)\, e^{it\omega} \right|^2 \{f(\omega)\}^{-1}\, d\omega, \tag{2}$$

where $$f(\omega) = (1 + \beta_1 e^{i\omega} + \ldots + \beta_h e^{ih\omega})(1 + \beta_1 e^{-i\omega} + \ldots + \beta_h e^{-ih\omega})\,\sigma^2/2\pi$$

is the spectral density of the process $\{x_t\}$. Also the second term in (2) may be replaced by

$$-\frac{n}{2\sigma^2} \sum_{s=-\phi(n)}^{\phi(n)} \alpha_s C_s,$$

where for $s \geqslant 0$, $C_s = \sum_{t=1}^{n-s} (x_t - \mu)(x_{t+s} - \mu)/(n-s)$ is the sample serial covariance for lag s, and $C_{-s} = C_s$, α_s is the coefficient of $e^{i\omega s}$ in the Fourier expansion of

$$\{g(\omega)\}^{-1} = \sigma^2\{2\pi f(\omega)\}^{-1} = (1 + \beta_1 e^{i\omega} + \ldots + \beta_h e^{ih\omega})^{-1}(1 + \beta_1 e^{-i\omega} + \ldots + \beta_h e^{-ih\omega})^{-1}, \tag{3}$$

and $\phi(n) \to \infty$, $n^{-1}\phi(n) \to 0$ as $n \to \infty$. Asymptotically efficient estimates $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_h$ are therefore obtained by minimizing $U = n \sum_s \alpha_s C_s$ with respect to $\beta_1, \beta_2, \ldots, \beta_h$. (If μ is not specified *a priori* it is easily seen that we need merely replace μ by $\bar{x} = \sum_{t=1}^{n} x_t/n$ in the definition of C_s.) These estimates are asymptotically normal, the limiting joint distribution of $\sqrt{n}(\hat{\beta}_i - \beta_i)$ $(i = 1, 2, \ldots, h)$ being multinormal with mean $\mathbf{O}$ and covariance matrix the inverse of the matrix whose (i, j)th element is the constant term in the Fourier expansion of

$$\frac{1}{2}\frac{\partial \log g(\omega)}{\partial \beta_i}\frac{\partial \log g(\omega)}{\partial \beta_j} \quad (i, j = 1, 2, \ldots, h).$$

Tests of hypotheses that certain of the β_i have specified values may also be made by means of the likelihood ratio principle, and the standard asymptotic theory for such tests is applicable.

The condition on the roots of $z^h+\beta_1 z^{h-1}+\ldots+\beta_h$ is incidentally not very restrictive since any stationary normal moving-average process has a 'regular' representation of the form (1) such that the moduli of the roots of $z^h+\beta_1 z^{h-1}+\ldots+\beta_h=0$ do not exceed unity (see, for example, Wold, 1954, p. 126). The only cases not covered are therefore those for which there is at least one root of modulus unity, arising when the spectral density of the process has at least one zero.

There are, however, various difficulties which may occur in using Whittle's method in practice. It might be thought that the obvious procedure for obtaining $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_h$ by solving the equations $\partial U/\partial\beta_i = 0$ $(i = 1, 2, \ldots, h)$ by successive approximation, taking as a first approximation the estimates obtained by equating the sample serial correlations $r_i = C_i/C_0$ $(i = 1, 2, \ldots, h)$ to the corresponding autocorrelations ρ_i of the process, would usually be satisfactory. However, the asymptotic efficiency of the first approximation will not be high unless $\beta_1, \beta_2, \ldots, \beta_h$ are small, so that the approximation process may converge slowly or perhaps even not at all; moreover, there may be an appreciable probability of the first approximation values not being real—this will happen if the polynomial $P(u)$ defined by $P(z+z^{-1}) \equiv \sum_{s=-h}^{h} r_s(z^s+z^{-s})$ has a zero with modulus less than 2 which is not of even multiplicity (Wold, 1954, p. 154). Also the expression of the α_s as explicit functions of $\beta_1, \beta_2, \ldots, \beta_h$, obtained by putting $\{g(\omega)\}^{-1}$ into partial fractions which are then expanded, will in general involve very tedious algebra. (The problem is the same as that of expressing the autocovariances of an autoregressive process in terms of its coefficients, which is considered by Quenouille (1947, pp. 365–6).) The results are quite complicated even when $h = 2$, for which

$$\{g(\omega)\}^{-1} = -\frac{A_2}{\beta_2} + \frac{z(A_1+A_2 z)}{1+\beta_1 z+\beta_2 z^2} + \frac{z^{-1}(A_1+A_2 z^{-1})}{1+\beta_1 z^{-1}+\beta_2 z^{-2}}$$

(where $z = e^{i\omega}$), whence

$$A_1 = -\frac{\beta_1}{(1-\beta_2)\{(1+\beta_2)^2-\beta_1^2\}}, \quad A_2 = \frac{-\beta_2(1+\beta_2)}{(1-\beta_2)\{(1+\beta_2)^2-\beta_1^2\}},$$

so that $\alpha_0 = -A_2/\beta_2, \quad \alpha_1 = \alpha_{-1} = A_1, \quad \alpha_2 = \alpha_{-2} = A_2 - \beta_1 A_1,$ etc.

The equations $\partial U/\partial\beta_i = 0$ will therefore be very cumbersome. Direct evaluation of the constant terms in the expansions of

$$\frac{1}{2}\frac{\partial \log g(\omega)}{\partial\beta_i}\frac{\partial \log g(\omega)}{\partial\beta_j},$$

which may be regarded as elements of the information matrix (per observation), is also troublesome, but can be avoided by using the fact that this information matrix is equal to the covariance matrix of h consecutive observations from a stationary autoregressive process $\{y_t\}$ such that

$$y_t+\beta_1 y_{t-1}+\ldots+\beta_h y_{t-h} = \xi_t,$$

where $\{\xi_t\}$ is a set of independent random variables with mean zero and variance unity. For the elements of the inverse of the latter covariance matrix (which is therefore equal to the asymptotic covariance matrix of $\sqrt{n}(\hat{\beta}_i-\beta_i)$ $(i = 1, 2, \ldots, h)$ are readily obtained in terms of $\beta_1, \beta_2, \ldots, \beta_h$ (see Durbin, 1959, p. 311, equation (19)).

With an electronic computer, one might determine the values $\hat{\beta}_i$ for which U is a minimum by evaluating U for a sufficiently large number of sets of values of the β_i, but even then the computations could be fairly laborious.

There is also what may be called the 'truncation' problem, the choice of the maximum lag $\phi(n)$ of the sample serial correlations occurring in the sum which gives U. Although the precise form of $\phi(n)$ does not affect the limiting distribution of the estimates $\hat{\beta}_i$, it will have a considerable influence on the accuracy of this limiting distribution as an approximation to the actual distribution for a given finite value of n. Presumably the more rapidly $\phi(n)$ increases with n, the more rapidly will the variances and covariances of the estimates approach their asymptotic values, but the more slowly will their biases decrease in magnitude, and their distribution tend to normality, so that one would have to strike a balance between these effects in seeking a $\phi(n)$ which was in some sense optimum. Determination of such an optimum $\phi(n)$, which in any case would depend on the true values of the parameters β_i, seems to be a formidable problem. In practice one might proceed by taking $\phi(n)$ to be such that $|\alpha_s^*/\alpha_0^*|$ is less than some prescribed small quantity, say 0·01, for all $s > \phi(n)$, where α_s^* denotes the value of α_s obtained by replacing the β_i by the first approximation to the $\hat{\beta}_i$ ($\alpha_s \to 0$ exponentially as $s \to \infty$). Alternatively one might determine a sequence of estimates $\hat{\beta}_i^{(k)}$ by minimizing $U_k = n \sum_{s=-k}^{k} \alpha_s C_s$ for values of k forming an increasing sequence $\{k_r\}$, and take $\phi(n)$ to be the least k_r such that the differences $|\hat{\beta}_i^{(k_r)} - \hat{\beta}_i^{(k_{r-1})}|$ are negligible. However, it is by no means clear that these two methods will always be satisfactory, and the second may require a very large amount of computation.

In this paper we present a modification of Whittle's method which enables these difficulties to be avoided to a large extent, and also usually requires much less computation. The idea on which this is based is extremely simple; it is merely a question of observing that since $U = nC_0 \sum_{s=-\phi(n)}^{\phi(n)} \alpha_s r_s$, estimates obtained by maximizing the likelihood of the first $\phi(n)$ sample correlations are equivalent to the asymptotically efficient estimates $\hat{\beta}_i$, and similarly that those obtained by maximizing the likelihood of the first k sample serial correlations are equivalent to the $\hat{\beta}_i^{(k)}$, for any finite value of k. An exact expression for the likelihood of a set of sample serial correlations is not known but the approximation derived from their asymptotic distribution (which is multinormal) can be used. The asymptotic behaviour of the resulting estimates based on $r_1, r_2, \ldots, r_k$ for any finite k is easily determined, and consideration of their asymptotic efficiencies will often enable one to deal satisfactorily with the truncation problem (see § 3).

An alternative method which enables asymptotically efficient estimates to be obtained with even less computation has been given by Durbin (1959). Here the moving-average process is approximated by an autoregressive process of order k, and the likelihood of the asymptotic distribution of the least-squares estimates of the coefficients of the autoregressive process (which are asymptotically equal to functions of $r_1, r_2, \ldots, r_k$) is then maximized. This method has the advantage of giving estimating equations which are linear in the unknown parameters, and can therefore be solved without iteration. However, there is the problem of determining the value of k (analogous to the truncation problem mentioned above), which Durbin does not discuss in detail, and this might sometimes make the application of the method difficult.

Another feature of the present method is that adjustments to the estimates to allow for their biases can be made fairly simply, as shown in § 4. With Durbin's method, on the other hand, the derivation of similar adjustments seems to be very tedious (the case of a first-order process is considered in the Appendix).

2. LARGE-SAMPLE ESTIMATES BASED ON A FINITE NUMBER OF SERIAL CORRELATIONS

Let k be a fixed positive integer, and L_k the logarithm of the likelihood of the asymptotic distribution of $r_1, r_2, \ldots, r_k$, which is to be maximized. L_k is most easily expressed as a function of the first h autocorrelations $\rho_1, \rho_2, \ldots, \rho_h$ of the moving-average process, and we therefore take these as the basic parameters rather than $\beta_1, \beta_2, \ldots, \beta_h$. The β_i are single-valued functions of the ρ_i, determined by

$$1 + \sum_{i=1}^{h} \beta_i z^{-i} = \prod_{i=1}^{h} (1 - z_i z^{-1}), \tag{4}$$

where $z_1, z_2, \ldots, z_h$ are the roots of the equation $1 + \sum_{i=1}^{h} \rho_i (z^i + z^{-i}) = 0$ whose moduli are less than unity. This is an immediate consequence of the identity

$$1 + \sum_{i=1}^{h} \rho_i (z^i + z^{-i}) \equiv \left(1 + \sum_{i=1}^{h} \beta_i z^i\right)\left(1 + \sum_{i=1}^{h} \beta_i z^{-i}\right) \Big/ \left(1 + \sum_{i=1}^{h} \beta_i^2\right).$$

Thus we shall obtain estimates $\hat{\rho}_i^{(k)}$ of ρ_i, and take the estimates $\hat{\beta}_i^{(k)}$ to be the same functions of the $\hat{\rho}_i^{(k)}$. The $\hat{\beta}_i^{(k)}$ of course cannot be determined if the equation $1 + \sum_{i=1}^{h} \hat{\rho}_i^{(k)} (z^i + z^{-i}) = 0$ has roots of the form $z = e^{i\phi}$ (where $0 \leqslant \phi \leqslant \pi$), but the probability of this happening will usually be small. (When it does happen we may take the corresponding estimates $\hat{z}_i^{(k)}$ to be substituted for z_i in (4) as $(e^{i\phi}, e^{-i\phi})$ repeated m times for a root of even multiplicity $2m$, and $+1$ or -1 repeated $2m+1$ times for a root of odd multiplicity $2m+1$, according as ϕ is nearer to 0 or π, except that if $\phi = 0$ or π, when the root must be of multiplicity $2m$, we repeat 1 or -1 m times.)

Now as $n \to \infty$, the joint distribution of $\sqrt{n}\,(r_i - \rho_i)$ $(i = 1, 2, \ldots, k)$ tends to the multinormal form with mean $\mathbf{O}$ and covariance matrix $W = (w_{ij})$, where

$$w_{ij} = \sum_{v=-\infty}^{\infty} \{\rho_v \rho_{v+i-j} + \rho_v \rho_{v+i+j} + 2\rho_i \rho_j \rho_v^2 - 2\rho_i \rho_v \rho_{v+j} - 2\rho_j \rho_v \rho_{v+i}\}, \tag{5}$$

ρ_v denoting the correlation between x_t and x_{t+v} for any integer v, so that $\rho_{-v} = \rho_v$, and $\rho_v = 0$ when $|v| > h$ (see, for example, Hannan, 1960, pp. 40–1). Hence

$$L_k = -\tfrac{1}{2} k \log 2\pi - \tfrac{1}{2} \log |W| - \frac{n}{2} \sum_{i,j=1}^{k} (r_i - \rho_i)\, w^{ij}\, (r_j - \rho_j), \tag{6}$$

where $W^{-1} = (w^{ij})$ is the inverse of W, so that

$$\frac{\partial L_k}{\partial \rho_s} = -\frac{1}{2|W|} \frac{\partial |W|}{\partial \rho_s} - \frac{n}{2}\left[\sum_{i,j=1}^{k} (r_i - \rho_i) \frac{\partial w^{ij}}{\partial \rho_s} (r_j - \rho_j) - 2 \sum_{i=1}^{h} w^{si} (r_i - \rho_i) - 2 \sum_{i=h+1}^{k} w^{si} r_i\right]$$
$$(s = 1, 2, \ldots, h). \tag{7}$$

Now let us assume that the equations obtained by putting these partial derivatives equal to zero have a consistent solution, i.e. one that converges in probability to $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_h)'$ when $n \to \infty$. The difference between this and $\boldsymbol{\rho}$ will then be of order $n^{-\frac{1}{2}}$ (as can be seen, for example, by applying the mean value theorem to (7)). It follows that an asymptotically equivalent solution

$$\hat{\boldsymbol{\rho}}^{(k)} = (\hat{\rho}_1^{(k)}, \hat{\rho}_2^{(k)}, \ldots, \hat{\rho}_h^{(k)})'$$

is obtained by neglecting the first two terms on the right-hand side of (7), giving

$$\sum_{i=1}^{h} \hat{w}^{si}(r_i - \hat{\rho}_i^{(k)}) = -\sum_{i=h+1}^{k} \hat{w}^{si} r_i \quad (s = 1, 2, \ldots, h),$$

where $\hat{w}^{si}$ is obtained from w^{si} by replacing $\boldsymbol{\rho}$ by $\hat{\boldsymbol{\rho}}^{(k)}$. Thus if $(\hat{u}_{ij})$ is the inverse of the matrix consisting of the estimated first h rows and columns of W^{-1}

$$\begin{aligned}\hat{\rho}_i^{(k)} &= r_i + \sum_{j=1}^{h} \hat{u}_{ij} \sum_{s=h+1}^{k} \hat{w}^{js} r_s \\ &= r_i + \sum_{s=h+1}^{k} \hat{c}_{is} r_s \quad \text{say} \quad (i = 1, 2, \ldots, h). \end{aligned} \tag{8}$$

These equations may be solved by iteration, the mth approximation

$$\boldsymbol{\rho}_m^{*(k)} = (\rho_{1m}^{*(k)}, \quad \rho_{2m}^{*(k)}, \quad \ldots, \quad \rho_{hm}^{*(k)})'$$

to $\hat{\boldsymbol{\rho}}$ being determined from the $(m-1)$th by

$$\rho_{im}^{*(k)} = r_i + \sum_{s=h+1}^{k} c_{is}(\boldsymbol{\rho}_{m-1}^{*(k)}) r_s,$$

and the first approximation being taken to be $\rho_{i1}^{*(k)} = r_i$. The values of the coefficients $c_{is}(\boldsymbol{\rho})$ for any particular $\boldsymbol{\rho}$ are most conveniently determined from the set of equations

$$\sum_{s=h+1}^{k} c_{is}(\boldsymbol{\rho}) w_{sj}(\boldsymbol{\rho}) = -w_{ij}(\boldsymbol{\rho}) \quad (j = h+1, \ldots, k) \tag{9}$$

which they clearly satisfy. It should be noted that the formula (5) may be rewritten

$$w_{ij} = \lambda_{i-j} + \lambda_{i+j} + 2(\rho_i \rho_j \lambda_0 - \rho_i \lambda_j - \rho_j \lambda_i), \tag{10}$$

where the λ_i are defined by

$$\left[1 + \sum_{i=1}^{h} \rho_i (z^i + z^{-i})\right]^2 \equiv \sum_{i=-2h}^{2h} \lambda_i z^i.$$

It is easily seen that $\hat{\boldsymbol{\rho}}^{(k)}$ is a consistent estimate of $\boldsymbol{\rho}$, and is asymptotically normal, the asymptotic distribution of $\sqrt{n}\,(\hat{\rho}_i^{(k)} - \rho_i)$ being the same as that of

$$\sqrt{n}\left(r_i - \rho_i + \sum_{s=h+1}^{k} c_{is}(\boldsymbol{\rho}) r_s\right),$$

provided that the functions c_{is} are bounded for all $\boldsymbol{\rho}$ which form a set of possible autocorrelations. That the c_{is} satisfy this condition follows from the equations (9), which show that they are rational functions of $\boldsymbol{\rho}$ with denominators which could be zero only if the determinant $|w_{sj}(\boldsymbol{\rho})|$ $(s, j = h+1, h+2, \ldots, k)$ were zero. (For $|w_{sj}(\boldsymbol{\rho})| = 0$ implies that some linear relation between the x_t holds with probability 1, which is clearly impossible for any $\boldsymbol{\rho}$.) The equations (9) moreover state that

$$r_i + \sum_{s=h+1}^{k} c_{is}(\boldsymbol{\rho}) r_s \quad (i \ldots 1, 2, \ldots, h)$$

are asymptotically uncorrelated with (and therefore independent of) r_j $(j = h+1, h+2, \ldots, k)$ so that we may write

$$r_i + \sum_{s=h+1}^{k} c_{is}(\boldsymbol{\rho}) r_s = r_i \cdot (r_{h+1}, r_{h+2}, \ldots, r_k),$$

the 'residual' obtained by subtracting from r_i its (asymptotic) regression on $r_{h+1}, r_{h+2}, \ldots, r_k$. Thus we have shown that $\hat{\rho}_i^{(k)}$ is asymptotically equivalent to $r_i.(r_{h+1}, r_{h+2}, \ldots, r_k)$. It should be noted that we have not needed to justify the assumption that the original maximum likelihood equations have a consistent solution, since we can suppose that the estimates are *defined* by the equations (8).

In applying the iteration method in practice it will usually not be worth while altering the arguments of the functions c_{is} after the first two or three iterations. In fact if $\tilde{\boldsymbol{\rho}}$ is any consistent estimate of $\boldsymbol{\rho}$ the asymptotic behaviour of the quantities

$$r_i + \sum_{s=h+1}^{k} c_{is}(\tilde{\boldsymbol{\rho}})\, r_s$$

is the same as that of the $\hat{\rho}_i^{(k)}$, and the chief reason for altering the arguments is to improve the accuracy of the asymptotic distribution as an approximation for finite n, one of the sources of error in the approximation being the deviation of the coefficients of r_s in the estimators from the corresponding coefficients in the true residual $r_i.(r_{h+1}, r_{h+2}, \ldots, r_k)$. Once the arguments of the c_{is} are fixed, the solution of the equations (9) can be easily obtained even when there is some doubt as to a suitable value of k. (The choice of k is discussed in § 3; clearly there may be a considerable loss of asymptotic efficiency if k is small.) For the method of triangular resolution of the matrix (w_{sj}), whereby a lower triangular matrix $L = (l_{ij})$ such that

$$w_{sj} = \sum_{i=h+1}^{j} l_{si} l_{ji} \quad (s \geqslant j = h+1, h+2, \ldots, k)$$

is determined, may be used, and an element l_{ij} depends only on the values of $w_{i'j'}$ for $i' \leqslant i$, $j' \leqslant j$, so that if it is desired that the value of k should be increased, the values of l_{ij} already computed do not need to be altered. (Compare Wold, 1949, who requires the triangular resolution of this same matrix.) It will often suffice to use a small initial value of k to determine the estimate, $\tilde{\boldsymbol{\rho}}$ say, which is to be substituted for $\boldsymbol{\rho}$ in the equations (9), giving the coefficients $w_{sj}(\tilde{\boldsymbol{\rho}})$, $w_{ij}(\tilde{\boldsymbol{\rho}})$ which are to be kept fixed for the remainder of the calculations.

As an illustration of the method, consider its application to one of the series of 100 observations used by Durbin (1959, § 8), which were generated by the model

$$x_t = \epsilon_t + \tfrac{1}{2}\epsilon_{t-1}.$$

For this series the first five sample serial correlations were:

$$r_1 = 0{\cdot}35005, \quad r_2 = -0{\cdot}06174, \quad r_3 = -0{\cdot}08007, \quad r_4 = -0{\cdot}14116, \quad r_5 = -0{\cdot}15629.$$

For a first-order process the matrix W is

$$\begin{pmatrix} 1-3\rho^2+4\rho^4 & 2(1-\rho^2) & \rho^2 & 0 & 0 & . \\ 2\rho(1-\rho^2) & 1+2\rho^2 & 2\rho & \rho^2 & 0 & . \\ \rho^2 & 2\rho & 1+2\rho^2 & 2\rho & \rho^2 & . \\ 0 & \rho^2 & 2\rho & 1+2\rho^2 & 2\rho & . \\ . & . & . & . & . & . \end{pmatrix} \tag{11}$$

(writing ρ for ρ_1 without ambiguity). Hence with $k = 2$, the first iteration gives

$$\rho_2^{*(2)} = r_1 - \left(\frac{2r_1(1-r_1^2)}{1+2r_1^2}\right) r_2 = 0{\cdot}38051.$$

The next iteration gives

$$\rho_3^{*(2)} = r_1 - \left(\frac{2\rho_2^{*(2)}(1-\{\rho_2^{*(2)}\}^2)}{1+2\{\rho_2^{*(2)}\}^2}\right) r_2 = 0{\cdot}38121.$$

This differs from $\rho_2^{*(2)}$ by a very small fraction of its asymptotic standard error, so that it does not matter which of the two estimates of ρ is taken as $\bar{\rho}$. (The value of the asymptotic standard error, calculated from formula (16) of § 3, is 0·053; this, incidentally, is little more than 10 % of the true value of 0·4 of the parameter ρ, so that the choice of $k = 2$ as an initial value is reasonable.)

With $\bar{\rho} = 0{\cdot}38051$, the equations (9) become

$$\begin{pmatrix} 1{\cdot}28958 & 0{\cdot}76102 & 0{\cdot}14479 & 0 & . \\ 0{\cdot}76102 & 1{\cdot}28958 & 0{\cdot}76102 & 0 & . \\ 0{\cdot}14479 & 0{\cdot}76102 & 1{\cdot}28958 & . & . \\ . & . & . & . & . \end{pmatrix} \times \mathbf{c} = \begin{pmatrix} 0{\cdot}65083 \\ 0{\cdot}14479 \\ 0 \\ . \end{pmatrix}.$$

The solutions of these for $k = 3, 4, 5$ are respectively

$$\mathbf{c}' = (-0{\cdot}67271, 0{\cdot}28471), \quad (-0{\cdot}73184, 0{\cdot}41599, -0{\cdot}16332),$$

and $$(-0{\cdot}75192, 0{\cdot}46547, -0{\cdot}24462, 0{\cdot}09210).$$

The corresponding estimates of ρ are

$$\hat{\rho}^{(3)} = 0{\cdot}36879, \quad \hat{\rho}^{(4)} = 0{\cdot}38498, \quad \hat{\rho}^{(5)} = 0{\cdot}37934.$$

3. Asymptotic efficiency of the estimates

Let $v_{ii}^{(k)}$ be the variance of the limiting distribution of $\sqrt{n}\,(\hat{\rho}_i^{(k)} - \rho_i)$. Since this limiting distribution is the same as that of $\sqrt{n}\, r_i \cdot (r_{h+1}, r_{h+2}, \ldots, r_k)$, we have

$$\begin{aligned} v_{ii}^{(k)} &= w_{ii} - \sum_{s=h+1}^{k} c_{is} w_{is} \\ &= w_{ii} - \sum_{s,\, s'=h+1}^{k} w_{is} w_{is'} a_{(k)}^{s's}, \end{aligned} \tag{12}$$

where the matrix $\{a_{(k)}^{s's}\}$ is the inverse of

$$A_{(k)} = \begin{pmatrix} w_{h+1,h+1} \cdots w_{h+1,k} \\ \vdots \\ w_{k,h+1} \quad \cdots \quad w_{kk} \end{pmatrix},$$

or

$$|A_{(k)}|\, v_{ii}^{(k)} = \left| \begin{array}{c|c} w_{ii} & w_{i,h+1} \cdots w_{ik} \\ \hline w_{h+1,i} & \\ \vdots & A_{(k)} \\ w_{ki} & \end{array} \right|. \tag{13}$$

Clearly $v_{ii}^{(k)}$ is a non-increasing function of k, $1 - \{v_{ii}^{(k)}/w_{ii}\}$ being the square of the asymptotic multiple correlation coefficient of r_i on $r_{h+1}, r_{h+2}, \ldots, r_k$, and therefore $\lim_{k\to\infty} v_{ii}^{(k)} = v_{ii}$ say, exists. In fact $n^{-1}v_{ii}$ $(i = 1, 2, \ldots, h)$ will be the asymptotic variances of asymptotically efficient estimates of $\rho_1, \rho_2, \ldots, \rho_h$. Similarly the asymptotic covariances of asymptotically efficient estimates will be $n^{-1}v_{ij}$ $(i \neq j)$, where

$$v_{ij} = \lim_{k\to\infty} v_{ij}^{(k)}$$

with

$$v_{ij}^{(k)} = w_{ij} - \sum_{s,\, s'=h+1}^{k} w_{is} w_{js'} a_{(k)}^{s's}, \tag{14}$$

the asymptotic covariance of $\sqrt{n}\,r_i.(r_{h+1}, r_{h+2}, \ldots, r_k)$ and $\sqrt{n}\,r_j.(r_{h+1}, r_{h+2}, \ldots, r_k)$. Explicit expressions for these asymptotic variances and covariances can be obtained by using a formula for the limits of elements of the variance-covariance matrix of consecutive observations from a moving-average process given by Walker (1961). For $w_{ij} = \lambda_{i-j}$ when $i, j > h$ (from equation (10)), so that $A_{(k)}$ is equal to the variance-covariance matrix of $k-h$ consecutive observations from a moving-average process whose covariance-generating function is

$$\left\{1 + \sum_{r=1}^{h} \rho_r (z^r + z^{-r})\right\}^2 = \left(1 + \sum_{r=1}^{h} \beta_r z^r\right)^2 \left(1 + \sum_{r=1}^{h} \beta_r z^{-r}\right)^2 \Big/ \left(1 + \sum_{r=1}^{h} \beta_r^2\right)^2.$$

Hence by the lemma in Walker (1961), if

$$A_{(k)}^{-1} = \{a_{(k)}^{h+i,h+j}\} \quad (i, j = 1, 2, \ldots, k-h),$$

$$a^{ij} = \lim_{k\to\infty} a_{(k)}^{h+i,h+j} = \left(1 + \sum_{r=1}^{h} \beta_r^2\right)^2 \sum_{r=1}^{\min(i,j)} \delta_{i-r}\delta_{j-r},$$

where

$$\left(1 + \sum_{r=1}^{h} \beta_r z^r\right)^{-2} \equiv \sum_{r=0}^{\infty} \delta_r z^r \quad [\text{for } |z| \leqslant 1]$$

or, taking $i \geqslant j$,

$$a^{ij} = \left(1 + \sum_{r=1}^{h} \beta_r^2\right)^2 \sum_{r=0}^{j-1} \delta_r \delta_{r+i-j}. \tag{15}$$

Also $w_{ij} = 0$ if $|i-j| > 2h$, so that from (12) and (14),

$$v_{ij}^{(k)} = w_{ij} - \sum_{s=h+1}^{2h+i} \sum_{s'=h+1}^{2h+j} w_{is} w_{js'} a_{(k)}^{s's} \quad \text{when } k > \max(i,j) + 2h, \tag{16}$$

and therefore

$$v_{ij} = w_{ij} - \sum_{r=1}^{h+i} \sum_{r'=1}^{h+j} w_{i,h+r} w_{j,h+r'} a^{r'r}. \tag{17}$$

It should be possible to verify that (v_{ij}) is the inverse of the information matrix for $\rho_1, \rho_2, \ldots, \rho_h$ given by Whittle's method. However, there seemed to be no way of obtaining manageable expressions for the elements of the information matrix, the constant terms in the Fourier expansions of

$$\frac{1}{2} \frac{\partial \log g(\omega)}{\partial \rho_i} \frac{\partial \log g(\omega)}{\partial \rho_j},$$

for an arbitrary value of h. When $h = 1$, the verification is easily performed. For then

$$a^{11} = (1+\beta^2)^2, \quad a^{12} = -2\beta(1+\beta^2)^2, \quad a^{22} = (1+4\beta^2)(1+\beta^2)^2,$$

so that

$$\begin{aligned} v_{11} &= w_{11} - (1+\beta^2)^2 \{w_{12}^2 - 4\beta w_{12} w_{13} + (1+4\beta^2) w_{13}^2\} \\ &= 1 - 3\rho^2 + 4\rho^4 - \rho^2(1+\beta^2)^2 \{4(1-\rho^2) - 8\rho\beta(1-\rho^2) + \rho^2(1+4\beta^2)\}, \end{aligned}$$

which with $\rho = \beta/(1+\beta^2)$ gives

$$v_{11} = (1-\beta^2)^3/(1+\beta^2)^4. \tag{18}$$

Also

$$\frac{\partial \log g(\omega)}{\partial \beta} = \frac{e^{i\omega}}{1+\beta e^{i\omega}} + \frac{e^{-i\omega}}{1+\beta e^{-i\omega}}$$

so that the constant term in the expansion of $\frac{1}{2}[\{\partial \log g(\omega)\}/\partial\beta]^2$ is $(1-\beta^2)^{-1}$ (cf. Whittle, 1953, p. 430). The constant term in the expansion of $\frac{1}{2}[\{\partial \log g(\omega)\}/\partial\rho]^2$ is therefore

$$\left(\frac{\partial\rho}{\partial\beta}\right)^{-2} (1-\beta^2)^{-1} = (1+\beta^2)^4/(1-\beta^2)^3,$$

the reciprocal of (18).

The asymptotic efficiency of $\hat{\rho}_i^{(k)}$ is thus equal to $v_{ii}/v_{ii}^{(k)}$. If an overall measure of asymptotic efficiency is required, the ratio $|v_{ij}|/|v_{ij}^{(k)}|$ may be used. When k is such that these asymptotic efficiencies are high (say not less than 0·80), little is to be gained by taking account of the information on $\boldsymbol{\rho}$ contained in the sample serial correlations for lags exceeding k, especially as the accuracy of the approximation of the asymptotic distribution of the estimates to their actual distribution for a given finite value of n may be expected to decrease as k increases. When prior information on $\boldsymbol{\rho}$ is available, this can be used to predict the value of k required to give estimates of high efficiency. In the absence of prior information, the prediction can be based on the estimate of $\boldsymbol{\rho}$ obtained with the small initial value of k. (In theory, of course, the value of k should not depend on the observations themselves, but it seems unlikely that this will matter in practice.)

To illustrate the variation of asymptotic efficiency with k and $\boldsymbol{\rho}$, a number of values of this were computed for the case $h = 1$, and are given in Table 1.

Table 1. *Asymptotic efficiency $E_k(\rho)$ of $\hat{\rho}^{(k)}$ for the first-order moving-average process*

$\lvert\rho\rvert$ \ k	1*	2	3	4
0·10	0·960	0·999	1·000	1·000
0·20	0·832	0·984	0·999	1·000
0·25	0·732	0·958	0·995	0·999
0·30	0·604	0·904	0·981	0·997
0·35	0·451	0·801	0·943	0·985
0·40	0·278	0·617	0·834	0·937
0·425	0·190	0·480	0·723	0·869
0·45	0·107	0·313	0·541	0·725

* For comparison the asymptotic efficiency $E_1(\rho)$ of r_1 is also given.

For a given value of k the asymptotic efficiency decreases as $|\rho|$ increases, as one would expect. As k increases the rate of decrease becomes very small except when $|\rho|$ is close to its largest possible value 0·5 (corresponding to $|\beta| = 1$), and the information on ρ contained in the sample serial correlations for lags exceeding 3 or 4 is negligible except for $|\rho| > 0{\cdot}4$. The main factor determining the value of $E_k(\rho)$ seems to be the magnitude of $|\beta|^{k+1}$, and the graphs of $E_k(\rho)$ plotted against $|\beta|^{k+1}$ for $k = 2, 3, 4$ are very similar (see Fig. 1). This is probably connected with the fact that the remainder after $k+1$ terms of the autoregressive representation

$$\sum_{i=0}^{\infty} (-\beta)^i x_{t-i} = \epsilon_t \tag{19}$$

is $-(-\beta)^{k+1}\epsilon_{t-k-1}$, whose standard deviation is $|\beta|^{k+1}\sigma$ (cf. Durbin, 1959, p. 307). If $|\beta|^{k+1} < 0{\cdot}05$, it should certainly be safe to conclude that the asymptotic efficiency is not less than 0·80.

In general the asymptotic efficiencies $E_{ki}(\boldsymbol{\rho}) = v_{ii}(\boldsymbol{\rho})/v_{ii}^{(k)}(\boldsymbol{\rho})$ will decrease as the moduli of the roots z_i of the equation $z^h + \beta_1 z^{h-1} + \ldots + \beta_h = 0$ increase. It is conjectured that for given k, the values of $E_{ki}(\boldsymbol{\rho})$ will depend mainly on the magnitude of the remainder after $k+1$ terms of the autoregressive representation which is the generalization of (19). It is easily seen that this remainder term can be written as

$$\sum_{i=1}^{h} \epsilon_{t-k-i}\left(\sum_{j=i}^{h} \beta_j \gamma_{i+k-j}\right), \tag{20}$$

where $\left(1+\sum_{r=1}^{h}\beta_r z^r\right)^{-1} \equiv \sum_{r=0}^{\infty}\gamma_r z^r$. The standard deviation of (20) is

$$\left\{\sum_{i=1}^{h}\left(\sum_{j=1}^{h}\beta_j\gamma_{i+k-j}\right)^2\right\}^{\frac{1}{2}},$$

and if this is small (say, less than 0·05), we should expect the $E_{ki}(\rho)$ to be near unity.

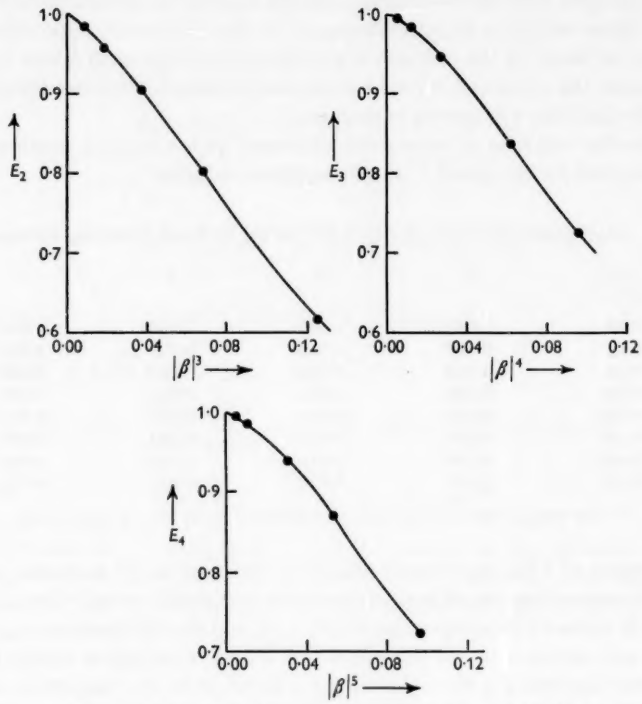

Fig. 1. Variation of Asymptotic Efficiency E_k with $|\beta|^{k+1}$

It should be noted that the formulae (12) and (14) at the beginning of this section will still apply when k is allowed to tend to infinity with n in a suitable way instead of being fixed. We then have estimates which are asymptotically efficient, equivalent to those obtained by Whittle's method. In this case a convenient computational procedure would be to calculate successive estimates

$$\boldsymbol{\rho}^{*(k)} = (\rho_1^{*(k)}, \rho_2^{*(k)}, \ldots, \rho_h^{*(k)})' \quad (k = 1, 2, 3, \ldots),$$

from the formula

$$\rho_i^{*(k)} = r_i + \sum_{s=h+1}^{k} c_{is}(\boldsymbol{\rho}^{*(k-1)})\, r_s, \tag{21}$$

starting with $\boldsymbol{\rho}^{*(1)} = (r_1, r_2, \ldots, r_h)'$. This procedure might in fact be regarded as a satisfactory substitute for the determination of estimates of $\beta_1, \beta_2, \ldots, \beta_h$ by minimizing $\sum_{s=-k}^{k}\alpha_s C_s$. Incidentally it is possible to verify from (21) that the distribution of the estimates $\rho_i^{*(k)}$ is

asymptotically normal by showing that the limiting distribution of $\sqrt{n}\,(\rho_i^{*(k)}-\rho_i)$ is the same as that of

$$\sqrt{n}\left\{r_i-\rho_i+\sum_{s=h+1}^{k} c_{is}(\boldsymbol{\rho})\,r_s\right\} \quad (i=1,2,\ldots,h),$$

the normality of which can be established by an extension of the argument used by Diananda (1953, pp. 240–1). However, the chief advantage of the present approach is that it enables one to consider any particular finite value of k, and so to avoid the difficulty of deciding on the rate at which k should tend to infinity.

4. Adjustments for biases of the estimates

It is well known that most large-sample estimates of parameters in time-series models which are functions of sample serial correlations (or have the same asymptotic distributions as such functions) are subject to biases which may not be negligible even for samples containing several hundred observations. This seems to be the case in the present problem but fortunately approximate allowances for these biases (which will be of order n^{-1}) can be made fairly easily for estimates equivalent to those given by the equations (8). To do so one uses the formula for the leading term in the expression for the bias of a sample serial correlation obtained by the usual series expansion method. Writing

$$n\{E(r_i)-\rho_i\}=\nu_i+o(1),$$

we have

$$\nu_i=2(\rho_i\lambda_0-\lambda_i) \tag{22}$$

when the process mean μ is specified *a priori*, and

$$\nu_i=2(\rho_i\lambda_0-\lambda_i)+(\rho_i-1)\sum_{s=-\infty}^{\infty}\rho_s \tag{22a}$$

when μ is not specified, and is replaced by the sample mean $\bar{x}$ (see, for example, Lomnicki & Zaremba, 1957, p. 156).

For

$$E\{\hat{\rho}_i^{(k)}\}-\rho_i=E\left\{\delta r_i+\sum_{s=h+1}^{k}(c_{is}(\boldsymbol{\rho})+\delta c_{is})\,\delta r_s\right\},$$

where

$$\delta r_i=r_i-\rho_i, \quad \delta c_{is}=c_{is}(\hat{\boldsymbol{\rho}}^{(k)})-c_{is}(\boldsymbol{\rho}),$$

which to order n^{-1} is equal to

$$n^{-1}\left\{\nu_i+\sum_{s=h+1}^{k}c_{is}(\boldsymbol{\rho})\,\nu_s\right\}+E\left(\sum_{s=h+1}^{k}\delta c_{is}\,\delta r_s\right).$$

Also

$$\delta c_{is}=\sum_{j=1}^{h}\frac{\partial c_{is}}{\partial\rho_j}(\hat{\rho}_j^{(k)}-\rho_j)=\sum_{j=1}^{h}\frac{\partial c_{is}}{\partial\rho_j}\left(r_j+\sum_{t=h+1}^{k}c_{jt}(\boldsymbol{\rho})\,r_t\right),$$

neglecting terms which are $o(n^{-\frac{1}{2}})$ (in probability), so that

$$E\left(\sum_{s=h+1}^{k}\delta c_{is}\,\delta r_s\right)=o(n^{-1})$$

since

$$r_j+\sum_{t=h+1}^{k}c_{jt}(\boldsymbol{\rho})\,r_t \quad (j=1,2,\ldots,h)$$

is asymptotically uncorrelated with r_s $(s=h+1,h+2,\ldots,k)$. Thus

$$E(\hat{\rho}_i^{(k)})-\rho_i=\left\{\nu_i+\sum_{s=h+1}^{k}c_{is}(\boldsymbol{\rho})\,\nu_s\right\}n^{-1}+o(n^{-1}). \tag{23}$$

An allowance for bias, to be subtracted from $\hat{\rho}_i^{(k)}$, will then be given by the estimate

$$\left\{\nu_i(\hat{\rho}^{(k)}) + \sum_{s=h+1}^{k} c_{is}(\hat{\rho}^{(k)})\,\nu_s(\hat{\rho}^{(k)})\right\} n^{-1} \tag{24}$$

of the right-hand side of equation (23). The adjusted estimates therefore become

$$\hat{\rho}_i^{(k)} - 2\left\{\hat{\rho}_i^{(k)}\lambda_0(\hat{\rho}^{(k)}) - \lambda_i(\hat{\rho}^{(k)}) - \sum_{s=h+1}^{k} c_{is}(\hat{\rho}^{(k)})\,\lambda_s(\hat{\rho}^{(k)})\right\} n^{-1} + \left[\sum_{s=-h}^{h} \hat{\rho}_s^{(k)}\left\{\hat{\rho}_i^{(k)} - 1 - \sum_{s=h+1}^{k} c_{is}(\hat{\rho}^{(k)})\right\}\right] \tag{25}$$

the term in the square bracket being included or omitted according as the ν_i are determined by (22a) or (22).

5. Numerical examples

The estimation process was applied to each of the twenty first-order series of 100 observations used by Durbin (1959) to illustrate his method, and referred to in § 2. The value $k = 5$ was used in equation (8), and the coefficients c_s were calculated using estimates of ρ obtained with the critical value $k = 2$, so that

$$\hat{\rho} = r_1 + \sum_{s=2}^{5} c_s(\tilde{\rho})\,r_s, \tag{26}$$

where

$$\tilde{\rho} = r_1 - \left(\frac{2r_1(1-r_1^2)}{1+2r_1^2}\right) r_2.$$

The effect of using an estimate $\tilde{\rho}$ based on more than the first two sample serial correlations was investigated for the first five of these series, and found in all cases to be a negligibly small fraction of the asymptotic standard error of $\hat{\rho}$. The values of $\hat{\rho}$ and $\hat{\beta} = \{1-(1-4\hat{\rho}^2)^{\frac{1}{2}}\}/2\hat{\rho}$ (the solution of the equation $\beta = \hat{\rho}(1+\beta^2)$ having modulus less than unity) are given in Table 2.

Table 2

Series	$\hat{\rho}$	$\hat{\beta}$	Series	$\hat{\rho}$	$\hat{\beta}$
1	0·37934	0·4594	11	0·40759	0·5162
2	0·33968	0·3918	12	0·40725	0·5155
3	0·33104	0·3785	13	0·41694	0·5373
4	0·40458	0·5097	14	0·40761	0·5162
5	0·39223	0·4842	15	0·26668	0·2889
6	0·42798	0·5642	16	0·36176	0·4280
7	0·41695	0·5373	17	0·34255	0·3964
8	0·38212	0·4646	18	0·44588	0·6140
9	0·40868	0·5186	19	0·35327	0·4138
10	0·34431	0·3992	20	0·41163	0·5252

The means and variances (with divisor 19) of these sets of values are

	$\hat{\rho}$	$\hat{\beta}$
Mean	0·38240	0·4730
Variance	0·001826	0·006042

The large-sample variance of $\hat{\rho}$ is given by

$$\begin{aligned} 100V(\hat{\rho}) &= v_{11}^{(5)}(0{\cdot}4) \\ &= 0{\cdot}6224 - (0{\cdot}7849 \times 0{\cdot}6720 - 0{\cdot}5110 \times 0{\cdot}1600) = 0{\cdot}1767 \end{aligned}$$

(using the formula at the beginning of § 3), and that of $\hat{\beta}$ by

$$100V(\hat{\beta}) = 100V(\hat{\rho})\left(\frac{\partial\rho}{\partial\beta}\right)^{-2}_{\beta=\frac{1}{2}} = 0{\cdot}1767 \times (1{\cdot}25)^4/(0{\cdot}75)^2 = 0{\cdot}7669.$$

(For $\beta = \frac{1}{2}$, the limiting value of v_{11} given by (18) is 0·1728, so that the asymptotic efficiency of $\hat{\rho}$ or $\hat{\beta}$ is 0·1728/0·1767 = 0·978.) The observed variances are in good agreement with these large-sample variances, but the deviations of the observed means from the theoretical values $\rho = 0{\cdot}4$, $\beta = 0{\cdot}5$ are rather large, the ratio of the deviation to its large-sample standard error being

$$-0{\cdot}0176/0{\cdot}00940 = -1{\cdot}88 \quad \text{for} \quad \bar{\hat{\rho}}$$

and

$$-0{\cdot}0270/0{\cdot}0196 = -1{\cdot}38 \quad \text{for} \quad \bar{\hat{\beta}}.$$

This suggests that the estimates are subject to negative biases which are not negligible. The allowances for bias of the estimates $\hat{\rho}$ given by the equation (24) will be approximately equal to the value of (23) when $\rho = 0{\cdot}4$ (with $i = 1$), i.e.

$$10^{-2}\{2(0{\cdot}4 \times 1{\cdot}32 - 0{\cdot}8 + 0{\cdot}785 \times 0{\cdot}16) + 1{\cdot}8(-0{\cdot}6 + 0{\cdot}446)\} = 0{\cdot}0057$$

(the values of the c_i when $\rho = 0{\cdot}4$ are

$$c_2 = 0{\cdot}7849, \quad c_3 = -0{\cdot}5110, \quad c_4 = 0{\cdot}2798, \quad c_5 = -0{\cdot}1076,$$

so that then $\Sigma c_i = 0{\cdot}446$).

This is rather more than half the large-sample standard error, and if we made the adjustment for bias suggested in § 4, the ratio of the deviation $\bar{\hat{\rho}} - 0{\cdot}4$ to its large-sample standard error would be reduced to approximately $-0{\cdot}0119/0{\cdot}0094 = -1{\cdot}3$. Similar allowances could be made for the biases of the estimates $\hat{\beta}$, using the approximation

$$\hat{\rho} - \rho \doteqdot (\hat{\beta} - \beta) f'(\beta) + \tfrac{1}{2}(\hat{\beta} - \beta)^2 f''(\beta),$$

where

$$f(\beta) = \beta/(1 + \beta^2),$$

so that

$$f'(\beta)\, E(\hat{\beta} - \beta) \doteqdot E(\hat{\rho} - \rho) - (1 - \beta^2) f''(\beta)/2n. \tag{27}$$

For $\beta = 0{\cdot}5$, $n = 100$, this gives $E(\hat{\beta} - \beta) \doteqdot -0{\cdot}020$, so that with these allowances the deviation $\bar{\hat{\beta}} - 0{\cdot}5$ would be only about 0·007, less than half the large-sample standard error.

It is interesting to compare these results with those obtained by Durbin's method (1959, § 8). His values of $\hat{\beta}$ are consistently smaller than the corresponding values in Table 2, their mean being 0·4531 compared with 0·4730, and the deviation of this from 0·5 is well over twice its large-sample standard error. This suggests that the present method of estimation is subject to less bias. It is, however, not at all clear why that should be so; certainly Durbin's method involves two estimation processes into which bias may enter (the estimation of the parameters of the approximating autoregressive process and then the estimation of β by a function of these estimated parameters), but so does the present method. A partial answer to the question can be given by obtaining an approximation, to order n^{-1}, to the bias of Durbin's estimates by a method similar to that used in § 4, and comparing this with the corresponding approximation to the bias of the present estimates derived from the right-hand side of equation (27). The approximation, which is derived in the Appendix, is, however, an extremely complicated function of β, and, because of the laborious computations required, the comparison was made only for $\beta = 0{\cdot}5$, $k = 5$. In this case, the approximation to the bias of Durbin's estimates becomes $-2{\cdot}2n^{-1}$ compared with $-2{\cdot}0n^{-1}$ for the bias of the present estimates, so that there seems to be little to choose between the two methods as regards bias. Hence it is probably fortuitous that for the twenty series of 100 observations the present method has given so much better results. It is incidentally interesting to note that the deviation of the mean of Durbin's twenty estimates from expectation is reduced to 0·025 when allowance is made for bias using the above approximation, the ratio of this to the standard error of the mean being only 1·3.

The approximation in the Appendix can of course be used to adjust Durbin's estimates for bias, but the complexity of the formula makes it somewhat inconvenient for practical application.

I wish to thank Mr J. Durbin and Miss J. May for providing me with the serial correlation coefficients for the twenty first-order series referred to in § 5, and also Mr D. A. East and Mr K. A. Tucker for undertaking most of the computations.

APPENDIX

Approximate adjustments for biases of Durbin's estimates of β for a first-order moving-average process

Let $\hat{\beta} = -\sum_{i=0}^{k-1} a_i a_{i+1} \Big/ \sum_{i=0}^{k} a_i^2$, where the a_i ($i > 0$) are the estimates of the coefficients α_i of the approximating autoregressive process (Durbin, 1959, equation (7)), and $a_0 = 1$. Putting $\delta a_i = a_i - \alpha_i$ and using

$$\beta\alpha_{i+1} + (1+\beta^2)\alpha_i + \beta\alpha_{i-1} = 0 \quad (i = 1, 2, \ldots, k),$$

with $\alpha_0 = 1$, $\alpha_{k+1} = 0$ (Durbin, 1959, p. 307), we obtain

$$\hat{\beta} = \left(\beta + (1-\beta^4)\beta^{-1}\sum_{i=0}^{k}\alpha_i\delta a_i - (1-\beta^2)\sum_{i=1}^{k-1}\delta a_i\delta a_{i+1}\right)\left(1 + 2(1-\beta^2)\sum_{i=1}^{k}\alpha_i\delta a_i + (1-\beta^2)\sum_{i=1}^{k}(\delta a_i)^2\right)^{-1}, \quad \text{(A 1)}$$

assuming k to be large enough for $\sum_{i=0}^{k-1}\alpha_i\alpha_{i+1}$ to be replaced by $\beta/(1-\beta^2)$ and $\sum_{i=0}^{k}\alpha_i^2$ by $1/(1-\beta^2)$.

On expanding the right-hand side of (A 1) and retaining only terms whose degree in the δa_i does not exceed 2, we find that to order n^{-1},

$$\hat{\beta} - \beta = (1-\beta^2)^2\beta^{-1}\sum_{i=1}^{k}\alpha_i\delta a_i - 2(1-\beta^2)^3\beta^{-1}\left(\sum_{i=1}^{k}\alpha_i\delta a_i\right)^2 - (1-\beta^2)\left(\beta\sum_{i=1}^{k}(\delta a_i)^2 + \sum_{i=1}^{k-1}\delta a_i\delta a_{i-1}\right). \quad \text{(A 2)}$$

Now if $\mathbf{a}' = (a_1, a_2, \ldots, a_k)$, $\mathbf{u}' = (r_1, r_2, \ldots, r_k)$, the r_i being the serial correlation coefficients defined in § 1, and $\mathbf{S} \equiv (s_{ij}) = (r_{i-j})$ $(i, j = 1, 2, \ldots, k)$ (with $r_0 = 1$), we have

$$\mathbf{a}'\mathbf{S} = -\mathbf{u}' \quad \text{or} \quad \mathbf{S}\mathbf{a} = -\mathbf{u}.$$

Let
$$\delta\mathbf{S} = \mathbf{S} - \boldsymbol{\Sigma}, \quad \text{where} \quad \boldsymbol{\Sigma} \equiv (\sigma_{ij}) = (\rho_{i-j}),$$
$$\delta\mathbf{a} = \mathbf{a} - \boldsymbol{\alpha} \quad \text{and} \quad \delta\mathbf{u}' = \mathbf{u}' - (\rho_1, \rho_2, \ldots, \rho_k).$$

Then
$$\delta\mathbf{S}\boldsymbol{\alpha} + \boldsymbol{\Sigma}\delta\boldsymbol{\alpha} + \delta\mathbf{S}\,\delta\mathbf{a} = -\delta\mathbf{u}, \quad \text{(A 3)}$$

assuming that the autocorrelations of the moving-average process can be taken to be equal to those of the approximating autoregressive process, so that $\boldsymbol{\alpha}'\boldsymbol{\Sigma} = -(\rho_1, \rho_2, \ldots, \rho_k)$. From (A 3),

$$\begin{aligned}\delta\mathbf{a} &= -\boldsymbol{\Sigma}^{-1}(\delta\mathbf{u} + \delta\mathbf{S}\boldsymbol{\alpha} + \delta\mathbf{S}\,\delta\mathbf{a}) \\ &= -\boldsymbol{\Sigma}^{-1}(\delta\mathbf{u} + \delta\mathbf{S}\boldsymbol{\alpha} - \delta\mathbf{S}\boldsymbol{\Sigma}^{-1}\{\delta\mathbf{u} + \delta\mathbf{S}\boldsymbol{\alpha}\}) + o(n^{-1}). \quad \text{(A 4)}\end{aligned}$$

Hence using the approximations

$$E(r_i - \rho_i) \doteq n^{-1}\nu_i \quad \text{(equation (22), § 4)},$$
$$E\{(r_i - \rho_i)(r_j - \rho_j)\} \doteq n^{-1}w_{ij},$$

where w_{ij} is defined by equation (5), § 1, and taking $w_{ij} = 0$ if either i or $j = 0$, we obtain the approximation

$$\begin{aligned}E(\delta a_i) &\doteq -n^{-1}\sum_{j=1}^{k}\sigma^{ij}\left\{\nu_j + \sum_{r=1}^{k}\nu_{j-r}\alpha_r - \sum_{r,s=1}^{k}\sigma^{rs}w_{|j-r|,s} - \sum_{r,s,t=1}^{k}w_{|j-r|,|t-s|}\sigma^{rt}\alpha_s\right\} \\ &= -n^{-1}\zeta_i, \quad \text{say.} \quad \text{(A 5)}\end{aligned}$$

Also using the expression for the asymptotic covariance matrix of the a_i due to Mann and Wald (see, for example, Durbin, 1959, p. 307 or Hannan, 1960, p. 50), we may take

$$E(\delta a_i\delta a_j) \doteq n^{-1}(1+\beta^2)^{-1}\sigma^{ij}. \quad \text{(A 6)}$$

Thus from (A 5), (A 6) and (A 2) an approximation to the bias of $\hat{\beta}$ is given by

$$-n^{-1}E(\hat{\beta}-\beta) \doteqdot (1-\beta^2)^2\beta^{-1}\sum_1^k \alpha_i\zeta_i + \left(\frac{1-\beta^2}{1+\beta^2}\right)\left\{2\beta^{-1}(1-\beta^2)^2\sum_{i,j=1}^k \alpha_i\alpha_j\sigma^{ij} + \beta\sum_{i=1}^k \sigma^{ii} + \sum_{i=1}^k \sigma^{i,i+1}\right\}. \quad \text{(A 7)}$$

The right-hand side of (A 7) with β replaced by $\hat{\beta}$ will then provide an adjustment for bias.

Computation of this right-hand side for $\beta = 0{\cdot}5$, $k = 5$, yielded the value 2·15. Hence for the series considered in § 5 the adjustments for bias to be applied to Durbin's estimates should be about 0·02.

It was suggested by Durbin that some improvement might be obtained by changing the denominator in $\hat{\beta}$ from $\sum_{i=0}^{k} a_i^2$ to $\sum_{i=0}^{k-1} a_i^2$. It is easily seen that since α_k is very small, the effect of this change on the approximation to the bias is very nearly equivalent to subtracting a term $\beta(1-\beta^2)(1+\beta^2)^{-1}\sigma^{kk}$ from the right-hand side of (A 7). For $\beta = 0{\cdot}5$, $k = 5$, this term is equal to 0·37, so that the approximate bias becomes $1{\cdot}78n^{-1}$, a reduction of nearly 20 %.

References

Diananda, P. H. (1953). Some probability limit theorems with statistical applications. *Proc. Camb. Phil. Soc.* **49**, 239–46.

Durbin, J. (1959). Efficient estimation of parameters in moving-average models. *Biometrika*, **46**, 306–16.

Hannan, E. J. (1960). *Time Series Analysis*. Methuen.

Lomnicki, Z. A. & Zaremba, S. K. (1957). On the estimation of autocorrelation in time series. *Ann. Math. Statist.* **28**, 140–58.

Quenouille, M. H. (1947). Notes on the calculation of autocorrelations of linear autoregressive schemes. *Biometrika*, **34**, 365–7.

Walker, A. M. (1961). On Durbin's formula for the limiting generalized variance of a sample of consecutive observations from a moving-average process. *Biometrika*, **48**, 197–9.

Whittle, P. 1951). *Hypothesis Testing in Time Series Analysis*. Uppsala: Almqvist and Wiksell.

Whittle, P. (1953). Estimation and information in stationary time series. *Ark. Mat. Astr. Fys.* **2**, 423–34.

Whittle, P. (1954). *A Study in the Analysis of Stationary Time Series, Appendix* 2. Uppsala: Almqvist and Wiksell.

Wold, H. (1949). A large sample test for moving-averages. *J. R. Statist. Soc.* B, **11**, 297–305.

Wold, H. (1954). *A Study in the Analysis of Stationary Time Series*. Uppsala: Almqvist and Wiksell.

Biometrika (1961), **48**, 3 *and* 4, *p.* 359
Printed in Great Britain

Linear and non-linear multiple comparisons in logit analysis

BY OLAV REIERSØL
University of Oslo, Norway

1. BERKSON'S MINIMUM LOGIT CHI-SQUARE METHOD DERIVED FROM NEYMAN'S REDUCED CHI-SQUARE AND LINEARIZATION METHOD

The minimum logit chi-square method of estimation has been developed and discussed by Berkson in a series of papers (Berkson, 1944, 1951, 1953, 1955). Taylor (1953) has shown that the estimators obtained by this method are regular best asymptotically normal (RBAN) estimators.* The same result can be obtained by showing that we may get Berkson's method of estimation as an application of Neyman's χ^2_1 and linearization method (Neyman, 1949). What Neyman denotes by χ^2_1 has been called the reduced chi-square by Fix, Hodges & Lehmann (1959).

We consider a sequence of independent experiments divided into s groups. Let n_j be the number of experiments in the jth group. In each of these n_j experiments a non-stochastic variable x is given the value x_j. In each experiment we note whether an event A occurs or not. In biological assay the variable x will be some function, for instance, the logarithm, of the dose or intensity of the stimulus applied to the experimental animal. The event A may, for instance, be the death of the animal. We suppose that the probability of A has the same value p_j for all experiments in the jth group. We suppose that

$$p_j = \frac{1}{1+e^{-(\alpha+\beta x_j)}}, \tag{1·1}$$

from which we get inversely

$$\alpha+\beta x_j = \log\frac{p_j}{1-p_j}. \tag{1·2}$$

Let q_j be the relative frequency of the event A in the jth group of experiments. Expanding $\log\{p_j/(1-p_j)\}$ in a Taylor series about $p_j = q_j$ and retaining only terms which are linear in p_j we get

$$\log\frac{p_j}{1-p_j} = \log\frac{q_j}{1-q_j}+\frac{1}{q_j(1-q_j)}(p_j-q_j). \tag{1·3}$$

Using (1·2) this may be written in the form

$$p_j-q_j = q_j(1-q_j)(\alpha+\beta x_j-u_j), \tag{1·4}$$

where

$$u_j = \log\frac{q_j}{1-q_j}.$$

Neyman's reduced chi-square is in this case

$$\sum_{j=1}^{s}\frac{n_j(p_j-q_j)^2}{q_j(1-q_j)}. \tag{1·5}$$

If we minimize (1·5) subject to (1·2) we get non-linear equations for the estimators. We shall, however, replace (1·2) by the linearized form (1·4). Substituting (1·4) in (1·5) we get

$$\sum_{j=1}^{s} n_j q_j(1-q_j)(\alpha+\beta x_j-u_j)^2. \tag{1·6}$$

* This concept was introduced by Neyman (1949) who used the term 'best asymptotically normal estimates'.

Neyman's reduced chi-square and linearization technique applied to the present case thus means that we shall use as estimators a and b those values of α and β which minimize (1·6). We have thus arrived at Berkson's minimum logit chi-square method.

It should be noted, however, that Berkson has recommended a modification of this method which consists in replacing each $q_j = 0$ by $q_j = 1/(2n_j)$ and each $q_j = 1$ by $q_j = 1 - 1/(2n_j)$ (Berkson, 1955, pp. 152–7).

This modification does not change the asymptotic properties of the estimators. The statements of the present paper will be valid whether we consider Berkson's original estimation method or the modified method just mentioned.

If we do not assume (1·1) *a priori*, we may regard it as an hypothesis to be tested. Neyman (1949) has presented several alternative tests each of which is equivalent in the limit to the likelihood ratio test. We shall consider the test criterion

$$\sum_j n_j q_j (1-q_j)(a+bx_j-u_j)^2 \tag{1·7}$$

which has asymptotically a chi-square distribution with $s-2$ degrees of freedom when the hypothesis (1·1) is true. A test based on the test criterion (1·7) is not a special case of any of Neyman's tests. It may, however, be shown that it is asymptotically equivalent to Neyman's tests.

2. A CHI-SQUARE TEST FOR DIFFERENCES BETWEEN SEQUENCES OF EXPERIMENTS

We shall consider k sequences $S_1, S_2, \ldots, S_k$, of experiments where each sequence is of the type considered in the preceding section. The sequence S_i is divided into s_i groups. In the jth of these groups x is given the value x_{ij} and we perform n_{ij} experiments in each of which the event A is supposed to occur with probability

$$p_{ij} = \frac{1}{1+e^{-(\alpha_i+\beta_i x_{ij})}}. \tag{2·1}$$

Let q_{ij} be the corresponding relative frequency. We then get RBAN estimators of the α_i and β_i by minimizing

$$\sum_{i=1}^{k}\sum_{j=1}^{s_i} w_{ij}(\alpha_i+\beta_i x_{ij}-u_{ij})^2, \tag{2·2}$$

where

$$w_{ij} = n_{ij}q_{ij}(1-q_{ij}), \quad u_{ij} = \log\frac{q_{ij}}{1-q_{ij}}.$$

u_{ij} is infinite when $q_{ij} = 0$ and when $q_{ij} = 1$. In both cases $w_{ij} = 0$. In such cases we may use the procedure advocated by Berkson and replace each $q_{ij} = 0$ by $q_{ij} = 1/(2n_{ij})$ and replace each $q_{ij} = 1$ by $q_{ij} = 1 - 1/(2n_{ij})$. Setting equal to zero the derivatives of (2·2) with respect to the α_i and β_i, we get the following equations for determination of the estimators a_i and b_i:

$$m_{0i}a_i + m_{1i}b_i = \sum_j w_{ij}u_{ij}, \tag{2·3}$$

$$m_{1i}a_i + m_{2i}b_i = \sum_j w_{ij}u_{ij}x_{ij}, \tag{2·4}$$

where

$$m_{0i} = \sum_j w_{ij}, \quad m_{1i} = \sum_j w_{ij}x_{ij}, \quad m_{2i} = \sum_j w_{ij}x_{ij}^2.$$

Let us consider the hypothesis

$$\alpha_1 = \alpha_2 = \ldots = \alpha_k = \alpha, \quad \beta_1 = \beta_2 = \ldots = \beta_k = \beta. \tag{2·5}$$

By analogy with the tests of Neyman (1949) we may use a test criterion which is the difference between the minimum of

$$\sum_{i,} w_{ij}(\alpha+\beta x_{ij}-u_{ij})^2 \tag{2·6}$$

with respect to α and β, and the minimum of (2·2) with respect to the α_i and β_i, which is

$$\sum_{i,j} w_{ij}(a_i+b_i x_{ij}-u_{ij})^2. \tag{2·7}$$

If we set $$m_0=\sum_{i,j} w_{ij},\quad m_1=\sum_{i,j} w_{ij}x_{ij},\quad m_2=\sum_{i,j} w_{ij}x_{ij}^2$$

the values a and b of α and β minimizing (2·6) will be determined by the equations

$$m_0 a+m_1 b=\sum_{i,j} w_{ij}u_{ij}, \tag{2·8}$$

$$m_1 a+m_2 b=\sum_{i,j} w_{ij}u_{ij}x_{ij} \tag{2·9}$$

and the minimum of (2·6) will be $$\sum_{i,j} w_{ij}(a+bx_{ij}-u_{ij})^2. \tag{2·10}$$

The difference between (2·10) and (2·7) is equal to

$$\sum_{i,j} w_{ij}(a_i-a+(b_i-b)x_{ij})^2,$$

which is again equal to

$$\sum_i \{m_{0i}(a_i-a)^2+2m_{1i}(a_i-a)(b_i-b)+m_{2i}(b_i-b)^2\}. \tag{2·11}$$

This test criterion has asymptotically a chi-square distribution with $2k-2$ degrees of freedom when the hypothesis (2·5) is true. Let us use the term 'accept' as a synonym of 'not reject'. Let $C_{\epsilon,k}$ be the upper ϵ-fractile of a chi-square distribution with k degrees of freedom. Then if we accept the hypothesis (2·5) when

$$\sum_i \{m_{0i}(a_i-a)^2+2m_{1i}(a_i-b)(b_i-b)+m_{2i}(b_i-b)^2\} \leqslant C_{\epsilon,2k-2}, \tag{2·12}$$

there is a probability asymptotically equal to $1-\epsilon$ that we shall accept (2·5) when it is true. The test described is asymptotically equivalent to the likelihood ratio test.

Adding equations (2·3) for $i=1,2,\ldots,k$ and subtracting (2·8) from the result we get

$$\sum_i m_{0i}(a_i-a)+\sum_i m_{1i}(b_i-b)=0. \tag{2·13}$$

Similarly we get $$\sum_i m_{1i}(a_i-a)+\sum_i m_{2i}(b_i-b)=0. \tag{2·14}$$

3. Linear multiple comparisons

If we reject the hypothesis (2·5) we shall usually be interested in knowing which α's and β's are different. We may also wish to compare two p-values corresponding to the same value of x in two different sequences of experiments. We may thus wish to test the hypotheses

$$\beta_i=\beta_j, \tag{3·1, ij}$$

$$\alpha_i+\beta_i x=\alpha_j+\beta_j x. \tag{3·2, ijx}$$

We may do this by means of an extension of the Scheffé multiple test (Scheffé, 1953, 1959).

Let z be a real column vector with n elements and let B be a real symmetric matrix which is positive definite. The Scheffé multiple test is based on the equivalence between the single inequality

$$z'B^{-1}z \leqslant c^2 \tag{3·3}$$

and the set of all inequalities

$$|h'z| \leqslant c(h'Bh)^{\frac{1}{2}} \quad \text{for all real column vectors } h. \tag{3·4}$$

We shall first prove that (3·3) and (3·4) are equivalent when z, B, h, c are all constant. We shall give a purely algebraic proof of this equivalence.

Since B is positive definite, it may be written as a product $A'A$ where A is a real square matrix. Then $B^{-1} = A^{-1}(A^{-1})'$. If we set $v = (A^{-1})'z$ and $g = Ah$, the inequalities (3·3) and (3·4) can be rewritten in the form

$$v'v \leqslant c^2 \tag{3·5}$$

and

$$|g'v| \leqslant c(g'g)^{\frac{1}{2}} \quad \text{for all real column vectors } g. \tag{3·6}$$

We shall now show that (3·5) and (3·6) are equivalent. Suppose first that (3·5) holds. Using the inequality of Cauchy (usually called the inequality of Schwarz) we get

$$(g'v)^2 \leqslant (g'g)\,(v'v) \leqslant c^2(g'g)$$

which gives (3·6).

Suppose next that (3·6) holds. Then we may set in (3·6) $g = v$ and get

$$v'v \leqslant c(v'v)^{\frac{1}{2}}$$

from which (3·5) follows. The equivalence is thus proved.

Suppose next that a sample space M is given. Let z, B, and c be random variables on the space M, and let H be a family of random variables h on the space M. Applying the equivalence between (3·3) and (3·4) to each point X of M, we find that if the inequality (3·3) holds at X, then the set of inequalities

$$|h'z| \leqslant c(h'Bh)^{\frac{1}{2}} \text{ for all } h \text{ belonging to } H \tag{3·7}$$

holds at X. Hence the set of points

$$\{X\colon z'B^{-1}z \leqslant c^2\} \tag{3·8}$$

is contained in the set

$$\{X\colon |h'z| \leqslant c(h'Bh)^{\frac{1}{2}} \text{ for all } h \text{ belonging to } H\}. \tag{3·9}$$

If $\Pr\{A\}$ denotes the probability of the statement A we thus have

$$\Pr\{|h'z| \leqslant c(h'Bh)^{\frac{1}{2}} \text{ for all } h \text{ belonging to } H\} \geqslant \Pr\{z'B^{-1}z \leqslant c^2\}. \tag{3·10}$$

Suppose next that the family H is such that

$$\left.\begin{array}{c}\text{For any real vector } k \text{ with } n \text{ components and any point } X \text{ of } M \text{ there exists a}\\ \text{random vector } h \text{ of the family } H \text{ which takes the value } k \text{ at the point } X.\end{array}\right\} \tag{3·11}$$

Then (3·3) and (3·7) are equivalent at each point X of M, hence the sets (3·8) and (3·9) are identical, and we have

$$\Pr\{|h'z| \leqslant c(h'Bh)^{\frac{1}{2}} \text{ for all } h \text{ belonging to } H\} = \Pr\{z'B^{-1}z \leqslant c^2\}. \tag{3·12}$$

Evidently (3·12) still holds if (3·11) holds with M replaced by a subset of M of probability one.

Since a constant is a particular kind of random variable, the statements above are still valid if some of the elements of z, B, c and the vectors of H are random and some of them

constants. If H contains all constant vectors with n components, then (3·11) holds, hence also (3·12) holds.

What we have presented here represents an extension of the results of Scheffé in the following respects:

(i) We do not assume that B is proportional to the covariance matrix of z. Actually, in the applications considered in the present paper, B is not proportional to the covariance matrix of z.

(ii) If the elements of z are subject to linear restrictions, our proof is still valid, while in the procfs given by Scheffé the existence of the inverse of the covariance matrix implies that the elements of z are linearly independent.

(iii) In the presentation of Scheffé the vectors h are supposed to be non-random.

It should be noted that Scheffé (1959, pp. 273–4) has considered the case of a random matrix B.

Using the results presented above and setting

$$\begin{bmatrix} m^{0i} & m^{1i} \\ m^{1i} & m^{2i} \end{bmatrix} = \begin{bmatrix} m_{0i} & m_{1i} \\ m_{1i} & m_{2i} \end{bmatrix}^{-1}$$

we conclude that at each point of the sample space the inequality (2·12) is equivalent to the set of all inequalities

$$\left|\sum_i h_{1i}(a_i-a)+\sum_i h_{2i}(b_i-b)\right| \leqslant (C_{\epsilon,2k-2}\sum_i (m^{0i}h_{1i}^2+2m^{1i}h_{1i}h_{2i}+m^{2i}h_{2i}^2))^{\frac{1}{2}} \qquad (3\cdot13)$$

for all real values of $h_{11}, \ldots, h_{1k}, h_{21}, \ldots, h_{2k}$. Because of the linear relations (2·13) and (2·14) there is an infinity of different sets of coefficients h_{1i} and h_{2i} giving the same function of the a_i and b_i. We have, for instance, if k_1 and k_2 are arbitrary real numbers

$$\begin{aligned} b_1-b_2 = {} & (1+k_1m_{11}+k_2m_{21})(b_1-b)-(1-k_1m_{12}-k_2m_{22})(b_2-b) \\ & +\sum_{i=3}^{k}(k_1m_{1i}+k_2m_{2i})(b_i-b)+\sum_{i=1}^{k}(k_1m_{0i}+k_2m_{1i})(a_i-a). \end{aligned} \qquad (3\cdot14)$$

The corresponding sum on the right-hand side of (3·13) is

$$m^{21}+m^{22}+f(k_1,k_2), \qquad (3\cdot15)$$

where

$$f(k_1,k_2)=\sum_{i=1}^{k}(m_{0i}k_1^2+2m_{1i}k_1k_2+m_{2i}k_2^2) \qquad (3\cdot16)$$

is a positive definite quadratic form. We thus get a set of inequalities

$$|b_1-b_2| \leqslant [C_{\epsilon,2k-2}\{m^{21}+m^{22}+f(k_1,k_2)\}]^{\frac{1}{2}} \qquad (3\cdot17)$$

which belongs to the set (3·13). The set (3·17) is equivalent to the single inequality which we get from (3·17) when $f(k_1,k_2)$ is given its minimum value which is 0. The set of inequalities (3·17) is therefore equivalent to the single inequality

$$|b_1-b_2| \leqslant \{C_{\epsilon,2k-2}(m^{21}+m^{22})\}^{\frac{1}{2}}.$$

More generally we get as special cases of (3·13)

$$|b_i-b_j| \leqslant \{C_{\epsilon,2k-2}(m^{2i}+m^{2j})\}^{\frac{1}{2}}, \qquad (3\cdot18, ij)$$

$$|a_i+b_ix-(a_j+b_jx)| \leqslant \{C_{\epsilon,2k-2}(m^{0i}+2m^{1i}x+m^{2i}x^2+m^{0j}+2m^{1j}x+m^{2j}x^2)\}^{\frac{1}{2}}. \qquad (3\cdot19, ijx)$$

If for all values of i, j, and x we accept the hypothesis $(3{\cdot}1, ij)$ when $(3{\cdot}18, ij)$ holds and $(3{\cdot}2, ijx)$ when $(3{\cdot}19, ijx)$ holds then, if the hypothesis $(2{\cdot}5)$ is true, there is an asymptotic probability greater than or equal to $1-\epsilon$ that all the hypotheses $(3{\cdot}1, ij)$ and $(3{\cdot}2, ijx)$ will be correctly accepted for all values of i, j and x. This statement follows from an application of $(3{\cdot}10)$.

4. NON-LINEAR MULTIPLE COMPARISONS

We shall consider a particular kind of non-linear comparison which arises when we wish to compare the medians of tolerance distributions corresponding to different sequences of experiments (cf. Finney, 1947, Chapter 4), or more generally if we wish to compare fractiles of tolerance distributions. The value of x corresponding to a given probability p of A in the ith sequence is

$$\frac{1}{\beta_i}\left(\log\frac{p}{1-p}-\alpha_i\right) \quad \text{or} \quad \frac{1}{\beta_i}(y-\alpha_i)$$

if we set $y = \log\{p/(1-p)\}$. We consider the hypothesis

$$\frac{1}{\beta_i}(y-\alpha_i) = \frac{1}{\beta_j}(y-\alpha_j)$$

which may preferably be written in the form

$$(y-\alpha_i)\beta_j - (y-\alpha_j)\beta_i = 0. \qquad (4{\cdot}1, ijy)$$

In the testing of this hypothesis we shall use the test criterion

$$T_{ijy} = (y-a_i)b_j - (y-a_j)b_i. \qquad (4{\cdot}2, ijy)$$

For any real number t we have the identity

$$(y-a_i)b_j-(y-a_j)b_i = (ta_i+(1-t)a_j-y)(b_i-b_j)+(tb_i+(1-t)b_j)(a_j-a_i). \qquad (4{\cdot}3)$$

We know that the set of inequalities $(3{\cdot}13)$ follows from $(2{\cdot}12)$ also in the case when the h_{1i} and h_{2i} are random variables. Setting

$$-h_{1i} = h_{1j} = tl_i+(1-t)b_j, \quad h_{2i} = -h_{2j} = ta_i+(1-t)a_j-y,$$

$h_{1m} = h_{2m} = 0$ when m is different from i and j, we get from $(3{\cdot}13)$

$$|T_{ijy}| \leqslant (C_{\epsilon,2k-2}G_{ijy}(t))^{\frac{1}{2}}, \qquad (4{\cdot}4)$$

where

$$G_{ijy}(t) = (m^{0i}+m^{0j})(tb_i+(1-t)b_j)^2+(m^{2i}+m^{2j})(ta_i+(1-t)a_j-y)^2 - 2(m^{1i}+m^{1j})(tb_i+(1-t)b_j)(ta_i+(1-t)a_j-y).$$

When i, j and y are given, $(4{\cdot}4)$ represents a set of inequalities, one for each value of t. This set is equivalent to the single inequality

$$|T_{ijy}| \leqslant (C_{\epsilon,2k-2}\min_t G_{ijy}(t))^{\frac{1}{2}}. \qquad (4{\cdot}5, ijy)$$

Using again $(3{\cdot}10)$ we may make the following statement: If we accept the hypothesis $(4{\cdot}1, ijy)$ when $(4{\cdot}5, ijy)$ holds, and if we accept the hypotheses $(3{\cdot}1, ij)$ and $(3{\cdot}2, ijx)$ as described in the preceding section, then, if the hypothesis $(2{\cdot}5)$ is true, the asymptotic probability is greater than or equal to $1-\epsilon$ that we shall simultaneously accept the hypotheses $(4{\cdot}1, ijy)$, $(3{\cdot}1, ij)$ and $(3{\cdot}2, ijx)$ for all values of i, j, x, y.

5. Additional remarks

There has been some controversy about the relative merits of Berkson's minimum logit chi-square method and the maximum likelihood method in logit analysis. We may particularly mention a paper by Silverstone (1957) and an answer to this paper by Berkson (1960).

The methods of the present paper may also be applied if we use maximum likelihood estimators or another kind of RBAN estimators in the test criterion (2·11). The methods may similarly be applied in probit analysis if we insert the estimators in an approximate chi-square obtained by a linearization analogous to that used in the present paper.

References

Berkson, J. (1944). Application of the logistic function to bioassay. *J. Amer. Statist. Ass.* **39**, 357–65.

Berkson, J. (1951). Relative precision of minimum chi-square and maximum likelihood estimates of regression coefficients. *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 471–9. University of California Press.

Berkson, J. (1953). A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function. *J. Amer. Statist. Ass.* **48**, 565–99.

Berkson, J. (1955). Maximum likelihood and minimum chi-square estimates of the logistic function. *J. Amer. Statist. Ass.* **50**, 130–62.

Berkson, J. (1960). Problems recently discussed regarding estimating the logistic curve. *Bull. Inst. Int. Statist.* **37**, 3e, 207–11. Discussion of the paper, same journal, **37**, 1re Livraison, p. 61.

Finney, D. J. (1947). *Probit Analysis* (2nd edition, 1952). Cambridge University Press.

Fix, E., Hodges, J. L. & Lehmann, E. L. (1959). The restricted chi-square test. *Probability and Statistics. The Harald Cramér Volume*, pp. 92–107. New York: John Wiley and Sons, Inc.

Neyman, J. (1949). Contribution to the theory of the χ^2 test. *Proc. Berkeley Symposium on Mathematical Statistics and Probability*, pp. 239–73. University of California Press.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, **40**, 87–104.

Scheffé, H. (1959). *The Analysis of Variance*. New York: John Wiley and Sons, Inc.

Silverstone, H. (1957). Estimating the logistic curve. *J. Amer. Statist. Ass.* **52**, 567–77.

Taylor, W. F. (1953). Distance functions and regular best asymptotically normal estimates. *Ann. Math. Statist.* **24**, 85–92.

Biometrika (1961), **48**, 3 *and* 4, *p.* 367
Printed in Great Britain

Estimation of the normal population parameters given a type I censored sample*

By J. G. SAW
University College London

1. Introduction

Of a sample of size n taken from a normal population with mean μ and standard deviation σ, only those observations which lie numerically between pre-assigned values d_1 and d_2 ($> d_1$) are available for statistical purposes. We suppose that a typical experiment yields the result $x_1 < \ldots < x_r < d_1 < x_{r+1} < \ldots < x_s < d_2 < x_{s+1} < \ldots < x_n$ so that s-r observations are available to us and that it is required to estimate μ and σ. Of course r and s are themselves dependent random variables; in fact r and $(s$-$r)$ have a trinomial distribution. The maximum likelihood estimators are investigated but prove bothersome to handle computationally.

According to Gupta's (1952) definition the sample above is described as being type I censored. An alternative form of censoring, defined in terms of fixed values r' and s' is as follows: of an ordered sample $x_1 < x_2 < \ldots < x_n$ we observe (or make use of) only those variates x_i for which $r' < i \leqslant s'$. The problem of estimation given a type II censored sample has received much attention (Gupta, 1952; Sarhan & Greenberg, 1956; Plackett, 1958; Saw, 1958, 1959). It seems to have been assumed quite widely that having observed a type I sample we may deal with it as though we have planned, prior to experimentation, that exactly r observations should be missing from the lower end of the sample and n-s observations from the upper end. Be this as it may, in dealing with type I censored samples as if they were type II censored, we seem to be ignoring the wealth of information stored in the experimental realization that r observations fell below d_1 and that n-s observations fell above d_2 and, in fact, in this paper we use this information alone to propose very simple, efficient (and indeed, familiar) estimators for μ and σ.

If d_1 and d_2 are finite, then the sample is said to be type I censored above and below; if $d_1 = -\infty$ the sample is said to be type I censored above only. This latter situation is also considered in the following sections. Certain symmetries exist which may be observed if we multiply all observations by -1, when we see that type I censoring above at d_2 corresponds to type I censoring below at $-d_2$. Similarly a procedure which holds for a sample type I censored below at d_1 and above at d_2 will hold for a sample type I censored below at $-d_2$ and above at $-d_1$.

When n is finite there is a non-zero probability that $s = r$, so that no observations are available, and in this case it is not possible to give a point estimate for (μ, σ). In the finite sample case therefore a procedure has to be adopted should such an instance arise; for example, the experimenter may choose to augment his sample by taking a further n' observations or to continue sampling item by item until one observation falls within the range d_1 to d_2. This indeterminacy is of importance for small sample sizes in that the bias

* The research reported here was supported in part by the Department of United States Army and was carried out at the University of North Carolina, Department of Bio-Statistics.

and variance of the estimators will be affected according as to what procedure is adopted. In the work which follows, however, we will be concerned only with large sample results so that the probability of a 'null' sample may be conveniently held to be zero.

2. NOTATION

Let
$$f(x) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad F(x) = \int_{-\infty}^{x} f(u)\,du, \tag{2·1}$$

$$z(x) = \frac{1}{\sqrt{(2\pi)}} \exp\left(-\tfrac{1}{2}x^2\right), \quad P(x) = \int_{-\infty}^{x} z(u)\,du, \tag{2·2}$$

$$d_1 = \mu + \sigma t_1, \quad d_2 = \mu + \sigma t_2, \tag{2·3}$$

$$F(d_1) = P(t_1) = p_1 = 1 - q_1, \quad F(d_2) = P(t_2) = p_2 = 1 - q_2, \tag{2·4}$$

$$x_i = \mu + \sigma v_i \quad (i = r+1, \ldots, s). \tag{2·5}$$

The observed sample is $(x_r) < d_1 < x_{r+1} < \ldots < x_s < d_2 < (x_{s+1})$; let y_{r+1}, y_{r+2}, ..., y_s be a random arrangement of $x_{r+1} < x_{r+2} < \ldots < x_s$ and write

$$y_i = \mu + \sigma u_i \quad (i = r+1, \ldots, s). \tag{2·6}$$

We see that $\{v_i\}$ is an ordered set drawn from the standard normal population and satisfies $t_1 < v_{r+1} < \ldots < v_s < t_2$ and that $\{u_i\}$ is a randomization of $\{v_i\}$. We will write

$$\rho_1 = r/n, \quad \rho_2 = s/n. \tag{2·7}$$

It will be convenient to use
$$z_i = z(t_i) \quad (i = 1, 2) \tag{2·8}$$

and where no limits are given, Σ is used as an abbreviation for $\sum_{r+1}^{s}$.

3. PROBABILTY DISTRIBUTION FUNCTION OF THE VARIATES $\{u_i\}$ AND $\{v_i\}$

We have that
$$p(r, s) = \frac{n!}{r!\,(s-r)!\,(n-s)!} p_1^r (p_2 - p_1)^{s-r} (1 - p_2)^{n-s}, \tag{3·1}$$

from which we may obtain moments and product moments of r and s. Also

$$p(v_{r+1}, \ldots, v_s, r, s) = \frac{n!}{r!\,(n-s)!} z(v_{r+1}) \ldots z(v_s)\, p_1^r (1 - p_2)^{n-s}, \tag{3·2}$$

$$p(v_i, r, s) = \frac{n!\, p_1^r (1-p_2)^{n-s} z(v_i)}{r!\,(i-r-1)!\,(s-i)!\,(n-s)!} (P(v_i) - p_1)^{i-r-1} (p_2 - P(v_i))^{s-i}, \tag{3·3}$$

$$p(v_i | r, s) = \frac{(s-r)!}{(i-r-1)!\,(s-i)!} \frac{z(v_i)\,(P(v_i) - p_1)^{i-r-1} (p_2 - P(v_i))^{s-1}}{(p_2 - p_1)^{s-r}}. \tag{3·4}$$

Since $\{u_i\}$ is a randomization of $\{v_i\}$ $(i = r+1, \ldots, s)$

$$p(u_{r+1}, \ldots, u_s | r, s) = z(u_{r+1}) \ldots z(u_s)/(p_2 - p_1)^{s-r}, \tag{3·5}$$

$$p(u_i | r, s) = z(u_i)/(p_2 - p_1). \tag{3·6}$$

It is obvious from (3·4) and (3·6) that the variable u_i is much more simple to deal with than the variable v_i. We shall make use of this by dealing with functions which are

symmetrical in x_i (and therefore in v_i) for then we may transform to the variates y_i (and therefore u_i) for clearly

$$\sum_{r+1}^{s} x_i = \sum_{r+1}^{s} y_i, \quad \text{etc.}$$

4. Maximum likelihood estimates

Let $\hat{\mu}$, $\hat{\sigma}$ be the maximum likelihood estimates of μ and σ. Since

$$\log p(x_{r+1}, \ldots, x_s, r, s) = \log(n!/r!(n-s)!) - (s-r)\log\sigma\sqrt{(2\pi)} - \tfrac{1}{2}\Sigma(x_i-\mu)^2/\sigma^2 + r\log F(d_1) + (n-s)\log(1-F(d_2)), \quad (4{\cdot}1)$$

we have that

$$\frac{\partial}{\partial\mu}\log p = \frac{1}{\sigma^2}\Sigma(x_i-\mu) - \frac{rf(d_1)}{F(d_1)} + (n-s)\frac{f(d_2)}{1-F(d_2)}, \quad (4{\cdot}2)$$

$$\frac{\partial}{\partial\sigma}\log p = -\frac{s-r}{\sigma} + \frac{1}{\sigma^3}\Sigma(x_i-\mu)^2 - r\frac{d_1-\mu}{\sigma}\frac{f(d_1)}{F(d_1)} + (n-s)\frac{d_2-\mu}{\sigma}\frac{f(d_2)}{1-F(d_2)}. \quad (4{\cdot}3)$$

The likelihood equations are therefore

$$\frac{1}{\hat{\sigma}}\Sigma(x_i-\hat{\mu}) - r\frac{z(\hat{t}_1)}{P(\hat{t}_1)} + \frac{(n-s)z(\hat{t}_2)}{1-P(\hat{t}_2)} = 0, \quad (4{\cdot}4)$$

$$\frac{1}{\hat{\sigma}^2}\Sigma(x_i-\hat{\mu})^2 - \frac{r\hat{t}_1 z(\hat{t}_1)}{P(\hat{t}_1)} + (n-s)\frac{\hat{t}_2 z(\hat{t}_2)}{1-P(\hat{t}_2)} = (s-r), \quad (4{\cdot}5)$$

where

$$\hat{t}_j = (d_j-\hat{\mu})/\hat{\sigma} \quad (j = 1, 2).$$

We may evaluate the expected value of the second derivatives for

$$\mathscr{E}\Sigma(x_i-\mu)/\sigma = \mathscr{E}\Sigma v_i = \mathscr{E}\Sigma u_i = -n(z_2-z_1) \quad (4{\cdot}6)$$

and similarly

$$\mathscr{E}\Sigma(x_i-\mu)^2/\sigma^2 = n(p_2-p_1-t_2z_2+t_1z_1), \quad (4{\cdot}7)$$

so that we arrive at

$$-\mathscr{E}\frac{\partial^2}{\partial\mu^2}\log p = \frac{n}{\sigma^2}\left[\left(p_2+\frac{z_2^2}{q_2}-t_2z_2\right)-\left(p_1-\frac{z_1^2}{p_1}-t_1z_1\right)\right], \quad (4{\cdot}8)$$

$$+\mathscr{E}\frac{\partial^2}{\partial\mu\partial\sigma}\log p = \frac{n}{\sigma^2}\left[\left(z_2+t_2^2z_2-\frac{t_2z_2^2}{q_2}\right)-\left(z_1+t_1^2z_1+\frac{t_1z_1^2}{p_1}\right)\right], \quad (4{\cdot}9)$$

$$-\mathscr{E}\frac{\partial^2}{\partial\sigma^2}\log p = \frac{n}{\sigma^2}\left[\left(2p_2-(t_2+t_2^3)z_2+\frac{t_2^2z_2^2}{q_2}\right)-\left(2p_1-(t_1+t_1^3)z_1-\frac{t_1^2z_1^2}{p_1}\right)\right]. \quad (4{\cdot}10)$$

We may thus determine the variance-covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ for any values of p_1 and p_2; elements in this matrix are given in Table 1 for various values of p_1 and p_2 and it may be noticed that

(i) If $p_2 \geqslant 0{\cdot}50$, then var $(\hat{\mu})$ varies little with p_1 provided that p_1 remains less than $\frac{1}{2}$.

(ii) If $p_2 < 0{\cdot}50$ and p_2-p_1 is fixed, then var $(\hat{\mu})$ decreases quite rapidly as p_2 closes to $0{\cdot}50$ this being particularly the case when p_2-p_1 is small.

(iii) If p_2-p_1 is fixed, var $(\hat{\sigma})$ varies little and is minimal when $p_1 = 0$ or when $p_2 = 1$.

For fixed p_2-p_1 therefore we verify the already intuitive idea that the best estimates of σ will result when the tail observations are available and that the best estimates of μ will obtain when the available observations are in the centre of the distribution.

5. A LINEAR ESTIMATOR FOR σ SYMMETRICAL IN $x_{r+1}, \ldots, x_s$

It will be conceded that the solution of the simultaneous equations (4·4) and (4·5) in μ and σ present computational difficulties. In order to overcome this and to gain advantage from the simplicity of the joint distribution of the y_i we investigate a linear estimator for σ symmetrical in $x_{r+1}, \ldots, x_s$; in its most general form this will be

$$L(\sigma: d_1, d_2) = a_0(\rho)\frac{1}{s-r}\Sigma x_i + a_1(\rho)\,d_1 + a_2(\rho)\,d_2, \quad |d_i| < \infty, \tag{5·1}$$

where $a_j(\rho) = a_j(\rho_1, \rho_2)$ is a function of the observed ratios $\rho_1 = r/n$ and $\rho_2 = s/n$. We may write (5·1) as

$$L(\sigma: d_1, d_2) = \mu(a_0(\rho) + a_1(\rho) + a_2(\rho)) + \sigma\left(a_0(\rho)\frac{1}{s-r}\Sigma u_i + t_1 a_1(\rho) + t_2 a_2(\rho)\right), \tag{5·2}$$

so that

$$\mathscr{E}(L(\sigma: d_1, d_2)|r, s) = \mu(a_0(\rho) + a_1(\rho) + a_2(\rho)) + \sigma\left(t_1 a_1(\rho) + t_2 a_2(\rho) - a_0(\rho)\frac{z_2 - z_1}{p_2 - p_1}\right). \tag{5·3}$$

We take

$$a_0(\rho) + a_1(\rho) + a_2(\rho) \equiv 0, \tag{5·4}$$

then $L(\sigma: d_1, d_2)$ will be unbiased to first order in $1/n$ if we take $\bar{z}_i = z(\bar{t}_i)$ such that $P(\bar{t}_i) = \rho_i$ $(i = 1, 2)$ and

$$\bar{t}_1 a_1(\rho) + \bar{t}_2 a_2(\rho) - a_0(\rho)\frac{\bar{z}_2 - \bar{z}_1}{\rho_2 - \rho_1} \equiv 1. \tag{5·5}$$

Putting

$$\left.\begin{aligned} \phi_i &= \frac{\partial}{\partial \rho_i}\left[t_1 a_1(\rho) + t_2 a_2(\rho) - a_0(\rho)\frac{z_2 - z_1}{p_2 - p_1}\right]_{\rho = p} && (i = 1, 2), \\ B_i &= \frac{t_i z_i}{p_2 - p_1} + \frac{z_i(z_2 - z_1)}{p_2 - p_1} && (i = 1, 2), \\ C &= (z_2 - z_1)/(p_2 - p_1), \end{aligned}\right\} \tag{5·6}$$

then the mean square error (M.S.E.) of $L(\sigma: d_1, d_2)$ is given to first order by

$$\text{M.S.E.}\,(L(\sigma)/\sigma) = \frac{1}{n}\left[(p_1 q_1 \phi_1^2 + 2p_1 q_2 \phi_1 \phi_2 + p_2 q_2 \phi_2^2) + \frac{a_0^2(p)}{p_2 - p_1}(1 + B_1 - B_2)\right], \tag{5·7}$$

and this is minimized, subject to the conditions of (5·4) and (5·5), by taking

$$\left.\begin{aligned} (-1)^{i+1}\phi_i D &= \frac{z_1^2 z_2^2 (t_2 - t_1)}{z_i(p_2 - p_1)} + \begin{vmatrix} p_1 q_1 q_2/q_i & -z_1 & -z_1 t_1 \\ p_1 p_2 q_2/p_i & -z_2 & -z_2 t_2 \\ 0 & 1 + B_1 - B_2 & B_1 t_1 - B_2 t_2 - C \end{vmatrix} && (i = 1, 2), \\ (a_0 D)(1 + B_1 - B_2) &= (\phi_1 D) z_1 + (\phi_2 D) z_2, \\ D &= (a_0 D)(B_1 t_1 - B_2 t_2 - C) - (\phi_1 D) z_1 t_1 - (\phi_2 D) z_2 t_2. \end{aligned}\right\} \tag{5·8}$$

Thus ϕ_1, ϕ_2 and $a_0(p)$; then

$$a_1 = a_0 B_1 - \phi_1 z_1, \quad a_2 = a_0 B_2 - \phi_2 z_2. \tag{5·9}$$

The actual weights used are of course those obtained by replacing ρ_i for p_i; $\bar{z}_i$ for z_i; $\bar{t}_i$ for t_i in equations (5·8) and (5·9). The results simplify when $d_1 = -\infty$ (i.e. when $p_1 = 0$), for

$$L(\sigma: -\infty, d_2) = -\frac{1/s}{(\bar{t}_2 + \bar{z}_2/\rho_2)}\sum_1^s (x_i - d_2), \tag{5·10}$$

and
$$\text{M.S.E.}\,(L(\sigma\colon -\infty, d_2)) = \frac{\sigma^2(1-B)}{np_2(t_2+z_2/p_2)^2}[1+p_2^2q_2(1-B)/z_2^2], \tag{5·11}$$

with
$$B = \left[\frac{t_2z_2}{p_2}+\frac{z_2^2}{p_2^2}\right].$$

The computations involved in (5·8) and (5·9) are not worth the effort in terms of the very small efficiency increase over another, extremely simple estimator for σ which is to be described in the next section. However, in the case of singly censored samples ($d_1 = -\infty$ or $d_2 = +\infty$) we are committed to the use of $L(\sigma\colon -\infty, d_2)$ unless we voluntarily censor the data below. Values of the mean square error of $L(\sigma\colon -\infty, d_2)$ are to be found in the first column of Table 2.

6. The quantile estimator for σ

Consider the function
$$\lambda(\sigma) = a(\rho)\,(d_2-d_1), \tag{6·1}$$

where $a(\rho) = a(\rho_1, \rho_2)$. We may write this as
$$\lambda(\sigma) = \sigma a(\rho)\,(t_2-t_1) \tag{6·2}$$

and observe that $\lambda(\sigma)$ will be unbiased for σ if $a(\rho)$ is such that
$$\mathscr{E}a(\rho)\,(t_2-t_1) = 1. \tag{6·3}$$

If we write
$$a(\rho) = 1/(\bar{t}_2-\bar{t}_1) \quad (r \neq s), \tag{6·4}$$

then it is easily shown that the bias of $\lambda(\sigma)$ is of order $1/n$. Writing
$$a(\rho) = (\bar{t}_2-\bar{t}_1)^{-1} - \frac{(\bar{t}_2-\bar{t}_1)^{-3}}{2n}\left[\frac{2p_1q_1}{z_1^2}-\frac{4p_1q_2}{z_1z_2}+\frac{2p_2q_2}{z_2^2}+(t_2-t_1)\left(\frac{t_1p_1q_1}{z_1^2}-\frac{t_2p_2q_2}{z_2^2}\right)\right]_{p=\rho}, \tag{6·5}$$

then the bias of $\lambda(\sigma)$ is of order $1/n^2$. In either case
$$\text{var}\left[\frac{\lambda(\sigma)}{\sigma}\right] = \frac{1}{n(t_2-t_1)^2}\left[\frac{p_1q_1}{z_1^2}-\frac{2p_1q_2}{z_1z_2}+\frac{p_2q_2}{z_2^2}\right]+O(1/n^2). \tag{6·6}$$

The value of var $(\lambda(\sigma)/\sigma)$ has been computed for several combinations of p_1 and p_2 and the asymptotic efficiency relative to $\hat{\sigma}$ is given in Table 2, where it will be seen that, at least in the range considered, the efficiency of this simple estimator is astonishingly high.

7. A linear estimator for μ symmetrical in $x_{r+1}, \ldots, x_s$

We take as estimator
$$L(\mu\colon d_1, d_2) = e_0(\rho)\frac{1}{s-r}\Sigma x_i + d_1e_1(\rho)+d_2e_2(\rho), \tag{7·1}$$

where $e_j(\rho) = e_j(\rho_1, \rho_2)$ $(j = 0, 1, 2)$ is a function of $\rho_1 = r/n$ and $\rho_2 = s/n$. We may write (7·1) as
$$L(\mu\colon d_1, d_2) = \mu(e_0(\rho)+e_1(\rho)+e_2(\rho))+\sigma\left[e_0(\rho)\frac{1}{s-r}\Sigma u_i + t_1e_1(\rho)+t_2e_2(\rho)\right], \tag{7·2}$$

so that
$$\mathscr{E}(L(\mu\colon d_1, d_2)|r, s) = \mu(e_0(\rho)+e_1(\rho)+e_2(\rho))+\sigma\left[t_1e_1(\rho)+t_2e_2(\rho)-e_0(\rho)\frac{z_2-z_1}{p_2-p_1}\right]. \tag{7·3}$$

If we take
$$e_0(\rho)+e_1(\rho)+e_2(\rho)\equiv 1, \tag{7·4}$$
then $L(\mu\colon d_1, d_2)$ will be unbiased to order $1/n$ if
$$\bar{t}_1 e_1(\rho)+\bar{t}_2 e_2(\rho)-e_0(\rho)\frac{\bar{z}_2-\bar{z}_1}{\rho_2-\rho_1}=0. \tag{7·5}$$

Writing
$$\theta_i=\frac{\partial}{\partial\rho_i}\left[t_1e_1(\rho)+t_2e_2(\rho)-\frac{z_2-z_1}{p_2-p_1}e_0(\rho)\right], \tag{7·6}$$
then with B_i and C defined as in (5·6), we find that $L(\mu\colon d_1, d_2)$ has least mean square error when $e_0(\rho)$, $e_1(\rho)$ and $e_2(\rho)$ are determined from

$$\left.\begin{aligned}(-)^j\theta_jK&=\frac{z_1^2z_2^2t_1t_2(t_2-t_1)}{z_jt_j(p_2-p_1)}(1+B_1-B_2)+(B_1t_1-B_2t_2-C)\\&\quad\times\begin{vmatrix}p_1q_1q_2/q_j & -z_1 & -z_1t_1\\ p_1p_2q_2/p_j & -z_2 & -z_2t_2\\ 0 & 1+B_1+B_2 & B_1t_1+B_2t_2-C\end{vmatrix},\\(e_0K)(B_1t_1-B_2t_2-C)&=(\theta_1K)z_1t_1+(\theta_2K)z_2t_2,\\K&=(e_0K)(1+B_1-B_2)-(\theta_1K)z_1-(\theta_2K)z_2.\end{aligned}\right\} \tag{7·7}$$

Thus e_0, θ_1 and θ_2; then
$$e_1=e_0B_1-\theta_1z_1,\quad e_2=-e_0B_2-\theta_2z_2. \tag{7·8}$$

Lastly
$$\text{M.S.E.}\,(L(\mu\colon d_1,d_2))=\frac{\sigma^2}{n}\left[p_1q_1\theta_1^2+2p_1q_2\theta_1\theta_2+p_2q_2\theta_2^2+\frac{e_0^2(1+B_1-B_2)}{p_2-p_1}\right]. \tag{7·9}$$

If $d_1=-\infty$, then
$$L(\mu\colon -\infty,d_2)=\frac{1}{(\bar{t}_2+\bar{z}_2/\rho_2)}\left[\frac{\bar{t}_2}{s}\sum_1^s x_i+\frac{\bar{z}_2}{2}d_2\right] \tag{7·10}$$

and
$$\text{M.S.E.}\,(L(\mu\colon -\infty,d_2))=\frac{\sigma^2}{np_2(t_2+z_2/p_2)^2}\left[1+t_2q_2\left(t_2+\frac{z_2}{p_2}\right)^2+t_2^2\left(1-\frac{t_2z_2}{p_2}-\frac{z_2^2}{p_2^2}\right)\right]. \tag{7·11}$$

Again we look at estimators which are nearly as efficient but more tractable.

8. Alternative estimators for μ of high efficiency

We now propose three estimators for μ which require very little labour in regard to their computation. Defining
$$C=(z_2-z_1)/(p_2-p_1),\quad \bar{C}=(\bar{z}_2-\bar{z}_1)/(\rho_2-\rho_1), \tag{8·1}$$
then an estimator for μ will be
$$\lambda_1(\mu\colon d_1,d_2)=\frac{1}{\bar{t}_1+\bar{C}}\left(\frac{\bar{t}_1}{s-r}\Sigma x_i+\bar{C}d_1\right), \tag{8·2}$$
which we may write as
$$\lambda_1(\mu\colon d_1,d_2)=\mu+\frac{\sigma}{\bar{t}_1+\bar{C}}\left(\frac{\bar{t}_1}{s-r}\Sigma u_i+\bar{C}t_1\right), \tag{8·3}$$
so that
$$\mathscr{E}[\lambda_1(\mu\colon d_1,d_2)|r,s)]=\mu+\frac{\sigma}{\bar{t}_1+\bar{C}}(t_1\bar{C}-\bar{t}_1C) \tag{8·4}$$

and therefore the bias is of order σ/n. With B_1 and B_2 defined as before it is easily shown that, to order $1/n^2$

$$\text{M.S.E.}\,(\lambda_1(\mu\colon d_1, d_2)) = \frac{\sigma^2}{n(t_1+C)^2}\left[(t_1B_1-C)^2\frac{p_1q_1}{z_1^2} - 2t_1B_2(t_1B_1-C)\frac{p_1q_2}{z_1z_2} + (t_1B_2)^2\frac{p_2q_2}{z_2^2} + \frac{t_1^2}{p_2-p_1}(1+B_1-B_2)\right]. \quad (8{\cdot}5)$$

Alternatively we may use the estimator

$$\lambda_2(\mu\colon d_1, d_2) = \frac{1}{\bar{t}_2+\bar{C}}\left[\frac{\bar{t}_2}{s-r}\Sigma x_i + \bar{C}d_2\right]. \quad (8{\cdot}6)$$

Its M.S.E. may be obtained from equation (8·5) by replacing t_1 by $-t_2$ in the numerator. The third and simplest estimator is

$$\lambda_3(\mu\colon d_1, d_2) = \frac{\bar{t}_2d_1-\bar{t}_1d_2}{\bar{t}_2-\bar{t}_1} \quad (8{\cdot}7)$$

which we may write as $$\lambda_3(\mu\colon d_1, d_2) = \mu + \sigma(\bar{t}_2t_1-t_2\bar{t}_1)/(\bar{t}_2-\bar{t}_1), \quad (8{\cdot}8)$$

so that the bias is of order σ/n after taking the expectation over r and s. It is easily shown that, to first order

$$\text{M.S.E.}\,(\lambda_3(\mu\colon d_1, d_2)) = \frac{\sigma^2}{n}\left[\frac{(t_2z_2)^2p_1q_1 - 2t_1t_2z_1z_2p_1q_2 + (t_1z_1)^2p_2q_2}{(t_2-t_1)^2z_1^2z_2^2}\right]. \quad (8{\cdot}9)$$

Which of the three estimators has the least mean-square error is governed by the prevailing values of p_1 and p_2. We note that if $p_1 = 0$ we are committed to the use of $\lambda_2(\mu\colon -\infty, d_2)$; similarly if $p_2 = 1$ we are committed to the use of $\lambda_1(\mu\colon d_1, -\infty)$. Investigation over the whole region $0 \leqslant p_1 < p_2 \leqslant 1$ indicates that

λ_3 is to be preferred if $0 < p_1 < p_2 \leqslant 0{\cdot}50$ or $0{\cdot}50 \leqslant p_1 < p_2 < 1$,

λ_1 is to be preferred if $0 \leqslant p_1 \leqslant 0{\cdot}50$ and $0{\cdot}50 \leqslant p_2 \leqslant 1-p_1$,

λ_2 is to be preferred if $0 \leqslant p_1 \leqslant 0{\cdot}50$ and $1-p_1 \leqslant p_2 \leqslant 1$.

Values of the asymptotic efficiency of λ_1, λ_2 and λ_3 are given in Table 3.

9. Summary

It is seen from Table 2 that when the sample is censored above and below the estimator $\lambda(\sigma\colon d_1, d_2)$ is of high asymptotic relative efficiency, at least in the range

$$0{\cdot}20 \leqslant p_1 < p_2 \leqslant 0{\cdot}80.$$

The estimator as defined in (6·2) and (6·4) is biased of order $1/n$ though this may be reduced to order $1/n^2$ using (6·5) (and of course the bias may be reduced to any order in $1/n$ by a more accurate evaluation of the expected value of $\lambda(\sigma\colon d_1, d_2)$ than was thought necessary for the purpose of this paper).

Similarly, the estimators $\lambda_j(\mu\colon d_1, d_2)$ $(j = 1, 2, 3)$, may be used for doubly censored samples as estimators of μ. Which of these three estimators is to be preferred is governed by the prevailing values of p_1 and p_2 and will therefore be indicated by the experimental values of ρ_1 and ρ_2.

These estimators, simple in form and in evaluation, were suggested as an alternative to maximum likelihood estimation which necessitates the solution of two simultaneous

equations ((4·4) and (4·5)) by an iterative process (Cohen, 1957); or to the method wherein the type I censored sample is treated as though it were type II censored. Although this last method is probably acceptable for large samples, a few calculations for small samples seems to indicate that the bias will often be quite large, also the experimental variance is tedious to calculate since it is necessary to obtain the conditional variance in every situation $1 \leqslant r < s \leqslant n$ with the associated probability of (r, s) in order to obtain the overall variance. On the other hand the variances of $\lambda(\sigma: d_1, d_2)$ and $\lambda_j(\mu: d_1, d_2)$ are comparatively very easy to determine.

In the case where the sample is singly censored, for example, censored above at d_2, the estimator $L(\mu: -\infty, d_2)$ is highly efficient. The estimator $L(\sigma: -\infty, d_2)$ is not, however, especially attractive since the asymptotic efficiency can be low when $p_2 > 0{\cdot}50$. It seems likely that in this case it would be advantageous to consider a quadratic estimator for σ^2 rather than a linear estimator for σ.

There is the possibility in finite samples that $\rho_1 = 0$ or $\rho_2 = 1$. In this event the limiting form as $\rho_1 \to 0$ (or $\rho_2 \to 1$) of $L(\mu: d_1, d_2)$ or of $L(\sigma: d_1, d_2)$ should be used. There is considerable simplification, for example if $\rho_1 = 0$ then B_1, defined in (5·6), is replaced by zero and B_2 by $z_2(t_2 + z_2)/p_2$.

When, simultaneously, $\rho_1 = 0$ and $\rho_2 = 1$, we conveniently use the sample mean and standard deviation as estimators of μ and σ, respectively. It should be remembered that if d_1 and d_2 are finite, the probability that $\rho_1 = 0$ or $\rho_2 = 1$ tends to zero as n tends to infinity.

REFERENCES

COHEN, A. C. (1957). On the solution of estimating equations for truncated and censored samples from normal populations. *Biometrika*, **44**, 225–36.

GUPTA, A. K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika*, **39**, 260–73.

PLACKETT, R. L. (1958). Linear estimation from censored data. *Ann. Math. Statist.* **29**, 131–42.

SARHAN, A. E. & GREENBERG, B. G. (1958). Estimation of the location and scale parameters by order statistics from single and doubly censored samples. *Ann. Math. Statist.* **29**, 79–105.

SAW, J. G. (1958). Moments of samples moments of censored samples from a normal population. *Biometrika*, **45**, 211–21.

SAW, J. G. (1959). Estimation of the normal population parameters given a singly censored sample. *Biometrika*, **46**, 150–5.

Table 1. *The asymptotic values of* var $(\hat{\mu})$, var $(\hat{\sigma})$ *and* cov $(\hat{\mu}; \hat{\sigma})$ *in units of* σ^2/n

p_2 \ p_1	0·00	0·20	0·25	0·30	0·35	0·40	0·45
0·20	5·7804						
	3·5375						
	3·7173						
0·25	4·0238	11·7985					
	2·7605	18·8166					
	2·5493	13·7156					
0·30	3·0199	5·1858	8·0081				
	2·2480	8·9293	18·9056				
	1·8322	5·6302	10·9363				
0·35	2·3982	3·1897	3·7350	5·2734			
	1·8853	5·6864	8·9971	18·9693			
	1·3574	3·0861	4·4304	8·3459			
0·40	1·9908	2·3019	2·4418	2·6897	3·3677		
	1·6155	4·0942	5·7376	9·0431	19·0122		
	1·0259	1·8973	2·3767	3·2819	5·8814		
0·45	1·7128	1·8312	1·8653	1·9098	1·9811	2·1583	
	1·4071	3·1544	4·1317	5·7715	9·0737	19·0370	
	0·7853	1·2323	1·4147	1·6848	2·1695	3·4982	
0·50	1·5171	1·5565	1·5615	1·5654	1·5682	1·5699	1·5706
	1·2414	2·5378	3·1821	4·1559	5·7921	9·0897	19·0464
	0·6052	0·8208	0·8777	0·9389	1·0057	1·0791	1·1607
0·55	1·3761	1·3859	1·3859	1·3866	1·3901	1·4011	1·4351
	1·1066	2·1039	2·5582	3·1993	4·1696	5·8016	9·0942
	0·4674	0·5488	0·5467	0·5255	0·4683	0·3344	Zero
0·60	1·2727	1·2755	1·2784	1·2855	1·3006	1·3324	
	0·9948	1·7830	2·1186	2·5698	3·2075	4·1740	
	0·3598	0·3606	0·3294	0·2731	0·1750	Zero	
0·65	1·1958	1·2021	1·2103	1·2253	1·2523		
	0·9004	1·5365	1·7931	2·1257	2·5736		
	0·2747	0·2262	0·1805	0·1098	Zero		
0·70	1·1383	1·1525	1·1662	1·1887			
	0·8198	1·3415	1·5428	1·7964			
	0·2066	0·1278	0·0755	Zero			
0·75	1·0949	1·1188	1·1374				
	0·7499	1·1834	1·3444				
	0·1516	0·0548	Zero				
0·80	1·0623	1·0958					
	0·6887	1·0525					
	0·1069	Zero					

Entries for (p_1, p_2) and (q_2, q_1) are identical.

Table 2. (*a*) *Asymptotic efficiency relative to* $\hat{\sigma}$ *of* $L(\sigma: -\infty, d_2)$.
(*b*) *Asymptotic efficiency relative to* $\hat{\sigma}$ *of* $\lambda(\sigma: d_1, d_2)$

	(*a*)	(*b*)					
p_2 \ p_1	0·00	0·20	0·25	0·30	0·35	0·40	0·45
0·20	0·9256						
·25	·9141	0·9992					
·30	·9022	·9962	0·9994				
·35	·8944	·9927	·9980	0·9997			
·40	·8763	·9906	·9966	·9990	0·9999		
·45	·8619	·9885	·9954	·9985	·9997	1·0000	
·50	·8460	·9869	·9946	·9980	·9995	0·9999	1·0000
0·55	0·8264	0·9858	0·9940	0·9977	0·9993	0·9998	0·9999
·60	·8086	·9851	·9936	·9975	·9991	·9997	
·65	·7858	·9845	·9930	·9969	·9985		
·70	·7591	·9833	·9916	·9953			
·75	·7271	·9805	·9882				
·80	·6873	·9739					

Table 3. (*a*) *Asymptotic efficiency relative to* $\hat{\mu}$ *of* $L(\mu: -\infty, d_2)$.
(*b*) *Asymptotic efficiency relative to* $\hat{\mu}$ *of* $\lambda_3(\mu: d_1, d_2)$

	(*a*)	(*b*)					
p_2 \ p_1	0·00	0·20	0·25	0·30	0·35	0·40	0·45
0·20	0·9485						
·25	·9479	0·9997					
·30	·9491	·9992	·09998				
·35	·9519	·9994	·9998	0·9998			
·40	·9559	·9997	·9999	·9999	0·9999		
·45	·9606	·9961	·9984	·9993	·9998	0·9999	
·50	·9658	·9909	·9941	·9966	·9983	·9994	0·9999
0·55	0·9711	0·9822	0·9863	0·9912	0·9950	0·9976	0·9992
·60	·9764	·9646	·9748	·9830	·9894	·9944	
·65	·9813	·9447	·9597	·9719	·9818		
·70	·9857	·9207	·9411	·9580			
·75	·9896	·8923	·9189				
·80	·9929	·8589					

Table 3. (*c*) *Asymptotic efficiency of* $\lambda_1(\mu: d_1, d_2)$ *relative to* μ *in the region* $0{\cdot}20 \leqslant p_1 < 0{\cdot}50 < p_2 \leqslant 0{\cdot}80$

p_2 \ p_1	0·20	0·25	0·30	0·35	0·40	0·45
0·55	0·9987	0·9998	0·9989	0·9954	0·9808	0·9738
·60	·9990	·9967	·9884	·9766	0·9587	
·65	·9937	·9847	·9708	·9513		
·70	·9845	·9696	·9493			
·75	·9722	·9517				
·80	·9573					

Note regarding Tables 2 *and* 3

In Tables 2 (*b*) and 3 (*b*), entries for (p_1, p_2) and (q_2, q_1) are identical so that, for example, the efficiency of $\lambda(\sigma: d_1, d_2)$ for the case $p_1 = 0{\cdot}60$, $p_2 = 0{\cdot}75$ is 0·9966. This is also true of Table 1.

In Table 3 (*c*), the efficiency of $\lambda_1(\mu: d_1, d_2)$ for (p_1, p_2) is the same as that of $\lambda_2(\mu: -d_2, -d_1)$ for (q_2, q_1) so that, for example, the efficiency of $\lambda_2(\mu: d_1, d_2)$ for the case $p_1 = 0{\cdot}40$, $p_2 = 0{\cdot}65$ is 0·9766.

In Tables 2 (*a*) and 3 (*a*), the entry for $(0{\cdot}00, p_2)$ is the same as that for $(q_2, 1{\cdot}00)$ so that, for example, the efficiency of $L(\mu: d_1, -\infty)$ when $p_1 = 0{\cdot}40$ is 0·9764.

Biometrika (1961), **48**, 3 *and* 4, *p.* 379
With 1 *text-figure*
Printed in Great Britain

Some tests for outliers†

BY C. P. QUESENBERRY‡ AND H. A. DAVID
Virginia Polytechnic Institute

1. INTRODUCTION

A recent issue of *Technometrics* (Vol. 2, 1960, pp. 123–66) contained two papers and an accompanying discussion on the rejection and location of outlying observations. It was emphasized, particularly in the discussion, that there might be several ways of approaching the problem, which depended to a large extent on the object in view. One might, for instance, be primarily interested in pruning the observations in order to secure a more accurate analysis of what was left, e.g. to obtain the most reliable estimate of a mean. Or one might be particularly interested in identifying which were the genuinely exceptional observations, in order to create a new insight into the phenomena under study. In the first case the criterion of what was best might be the effect on the standard error of estimation, in the second case the risk of wrongly deciding whether an observation was exceptional or not. The procedures discussed in the following paper start from the basis of risks of misclassification rather than of estimation errors.

The particular problem which we take is that of detecting outlying observations when in addition to the normal sample $x_1, x_2, \ldots, x_n$ at hand an independent mean-square estimate s_ν^2 of the common variance σ^2 is available. The same situation has been considered by Nair (1948). To test for one outlier at a specified end of the sample he proposes using the ratio of the extreme deviate from the sample mean to s_ν. For two-sided testing the extreme absolute deviate from the sample mean divided by s_ν has been proposed by Halperin, Greenhouse, Cornfield & Zalokar (1955).

The two statistics mentioned above do not make use of the variance estimate s^2 from the sample and for this reason do not possess certain desirable optimal properties. We shall propose statistics with the same numerators as those above but with s_ν in the denominators replaced by the pooled estimate

$$s^* = \{[(n-1)s^2 + \nu s_\nu^2]/(n+\nu-1)\}^{\frac{1}{2}}.$$

Kudô (1956) has shown that among a suitably restricted class of tests these statistics maximize the probability of rejecting the null hypothesis of homogeneity of the sample in the presence of a single outlier.

We shall develop a method for computing percentage points of these statistics and present the computed tables. These tables are also immediately applicable to the problem of slippage of means in normal samples. Two examples illustrate the procedures.

† This research supported in part by a National Science Foundation Fellowship and in part by the Office of Ordnance Research, U.S. Army.

‡ Now at Montana State College.

2. NOTATION AND DEFINITIONS

Let $x_1, x_2, \ldots, x_n$ denote the sample in the order drawn. Then define

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

s_ν^2 is an independent mean-square estimate of variance, σ^2, with ν degrees of freedom,

$$S^2 = (n-1)s^2 + \nu s_\nu^2,$$

$$b_i = (x_i-\bar{x})/S, \tag{2·1}$$

$$b = \max_i b_i = \frac{x_{\max}-\bar{x}}{S}, \tag{2·2}$$

$$b^* = \max_i |b_i|. \tag{2·3}$$

Then b and b^* are essentially the one-sided and two-sided statistics, respectively, discussed in § 1. It should be noted that S^2 has not been divided by its degrees of freedom.

The special case $\nu = 0$ has been treated by Pearson & Chandrasekhar (1936), Grubbs (1950) and Borenius (1958).

3. DISTRIBUTION THEORY

For the work in later sections the distribution of b_i and the joint distribution of b_i and b_j are needed. We shall now obtain these distributions, taking for definiteness $i = 1$ and $j = 2$. An extremely complicated derivation of essentially the same distributions has been given by Doornbos, Kesten & Prins (1956) in an article concerned with slippage tests.

As is well known, $(n-1)s^2$ may be decomposed into two independent components

$$\frac{n}{n-1}(x_1-\bar{x})^2 + \chi^2_{(n-2)}\sigma^2,$$

which are distributed respectively as $\chi^2\sigma^2$ with 1 and $n-2$ degrees of freedom. With the same notation we have, therefore

$$S^2 = \frac{n}{n-1}(x_1-\bar{x})^2 + \chi^2_{(n+\nu-2)}\sigma^2. \tag{3·1}$$

Then b_1 may be written as

$$b_1 = \left(\frac{n}{n-1} + \frac{\chi^2_{(n+\nu-2)}\sigma^2}{(x_1-\bar{x})^2}\right)^{-\frac{1}{2}}$$

$$= \left(\frac{n-1}{n}\right)^{\frac{1}{2}}\left(1+\frac{n+\nu-2}{t^2_{(n+\nu-2)}}\right)^{-\frac{1}{2}},$$

where $t_{(n+\nu-2)}$ denotes a t-variate with $n+\nu-2$ degrees of freedom. It follows that the density function of b_1 is

$$f(b_1) = \left(\frac{n}{n-1}\right)^{\frac{1}{2}} \frac{\Gamma[\frac{1}{2}(n+\nu-1)]}{\sqrt{\pi}\,\Gamma[\frac{1}{2}(n+\nu-2)]}[1-nb_1^2/(n-1)]^{\frac{1}{2}(n+\nu-4)}, \quad -\left(\frac{n-1}{n}\right)^{\frac{1}{2}} \leqslant b_1 \leqslant \left(\frac{n-1}{n}\right)^{\frac{1}{2}}. \tag{3·2}$$

This generalization of a result due to Thompson (1935) has also recently been pointed out by Anscombe (1960).

Continuing the decomposition of (3·1) one step further we have

$$S^2 = \frac{n}{n-1}(x_1-\bar{x})^2+\frac{n-1}{n-2}(x_2-\bar{x}')^2+\chi^2_{(n+\nu-3)}\sigma^2$$

with $$\bar{x}' = \frac{1}{n-1}\sum_{i=2}^{n} x_i.$$

Let $$b_2' = (x_2-\bar{x}')/S',$$

where $$S'^2 = S^2-\frac{n}{n-1}(x_1-\bar{x})^2.$$

Clearly, the distribution of b_2' is of the form (3·2) with n replaced by $n-1$. Moreover, b_2' is independent of b_1. To see this suppose s_ν^2 is based on a random sample of size $\nu+1$, with mean $\bar{x}_1$, taken from a $N(\mu_1, \sigma^2)$ parent. This assumption is unnecessarily restrictive but does not essentially affect the argument. Then $\bar{x}$, $\bar{x}_1$ and S^2 are complete sufficient statistics for μ, μ_1 and σ^2. Since the distribution of b_2' does not involve these parameters, b_2' is independent of the joint distribution of $\bar{x}$, $\bar{x}_1$ and S (Basu, 1955). Also b_2' does not involve x_1 so that it must be independent of b_1.

The joint density function of b_1 and b_2' is therefore

$$f(b_1, b_2') = \left(\frac{n}{n-2}\right)^{\frac{1}{2}}\left(\frac{n+\nu-3}{2\pi}\right)\left(1-\frac{n}{n-1}b_1^2\right)^{\frac{1}{2}(n+\nu-4)}\left(1-\frac{n-1}{n-2}b_2'^2\right)^{\frac{1}{2}(n+\nu-5)},$$

$$-\left(\frac{n-1}{n}\right)^{\frac{1}{2}} \leqslant b_1 \leqslant \left(\frac{n-1}{n}\right)^{\frac{1}{2}}, \quad -\left(\frac{n-2}{n-1}\right)^{\frac{1}{2}} \leqslant b_2' \leqslant \left(\frac{n-2}{n-1}\right)^{\frac{1}{2}}.$$

Since $$b_2' = \frac{b_2+b_1/(n-1)}{[1-nb_1^2/(n-1)]^{\frac{1}{2}}}$$

we obtain $$f(b_1, b_2) = \left(\frac{n}{n-2}\right)^{\frac{1}{2}}\frac{n+\nu-3}{2\pi}\left(1-\frac{n-1}{n-2}b_1^2-\frac{2b_1b_2}{n-2}-\frac{n-1}{n-2}b_2^2\right)^{\frac{1}{2}(n+\nu-5)} \tag{3·3}$$

over the ellipse $$\frac{n-1}{n-2}b_1^2-\frac{2b_1b_2}{n-2}+\frac{n-1}{n-2}b_2^2 \leqslant 1$$

$$= 0 \quad \text{elsewhere.}$$

Moments of b and b^*

Since the distributions of b and b^* do not involve μ, μ_1, and σ^2 it follows as above that b and b^* are distributed independently of S. This result is well known for the special case $\nu = 0$ and may indeed be proved in a similar fashion, as was pointed out to us by Dr G. E. P. Box. As a consequence of this independence the moments (about zero) of b and b^* are the ratios of the moments of their respective numerators and denominators. Thus we have, for example,

$$\mathscr{E}(b^r) = \frac{\mathscr{E}(x_{\max}-\bar{x})^r}{\mathscr{E}(S^r)}.$$

In this case (but not so readily for b^*) the right-hand side can be evaluated numerically since the cumulants of $x_{\max}-\bar{x}$ are related to the tabulated cumulants of the extreme Rub n, 1954) by equations which for $\mu = 0$, $\sigma = 1$ become (McKay, 1935)

$$\kappa_r(x_{\max}-\bar{x}) = \kappa_r(x_{\max}) \quad (r = 1, 3, 4, 5, \ldots),$$

$$\kappa_2(x_{\max}-\bar{x}) = \kappa_2(x_{\max})-1/n.$$

The distribution of b may therefore be approximated by a Pearson-type curve. However, for the purpose of obtaining upper percentage points the approach of the following section, applicable to both b and b^*, is preferable.

4. THE COMPUTATIONAL PROCEDURES

(a) *One-sided case, b*

We now consider a procedure for computing significance points of b defined by (2·2). For a given value of n and ν let D_α be the required α significance point of b. By Bonferroni's inequalities (cf. David, 1956) we have for any D

$$n\Pr(b_i > D) - \binom{n}{2}\Pr(b_i > D, b_j > D) \leqslant \Pr(b > D) \leqslant n\Pr(b_i > D). \tag{4·1}$$

For sufficiently large D the right side serves as a first approximation to $\Pr(b > D)$ and the left side as a second approximation. If the first approximation is set equal to α, then the resulting equation, i.e.

$$\Pr(b_i > D) = \alpha/n$$

can be solved for a value D_1, which is an upper bound of the value D_α sought. From (3·2) this equation can be written as

$$\alpha/n = C_1 \int_{D_1}^{\{(n-1)/n\}^{\frac{1}{2}}} \left[1 - \frac{n}{n-1} b_1^2\right]^{\frac{1}{2}(n+\nu-4)} db_1,$$

where

$$C_1 = \left[\frac{n}{\pi(n-1)}\right]^{\frac{1}{2}} \frac{\Gamma[\frac{1}{2}(n+\nu-1)]}{\Gamma[\frac{1}{2}(n+\nu-2)]}.$$

The following equivalent equation is more convenient to work with:

$$\begin{aligned} B_1 &= \int_0^{D_1} \left[1 - \frac{n}{n-1} x^2\right]^{\frac{1}{2}(n+\nu-4)} dx \\ &= D_1 + \sum_{r=1}^{\infty} \frac{(-1)^r [n+\nu-4][n+\nu-6] \ldots [n+\nu-2(r+1)] n^r D_1^{2r+1}}{r!\, 2^r (n-1)^r (2r+1)}, \end{aligned} \tag{4·2}$$

where

$$B_1 = \frac{\frac{1}{2} - \alpha/n}{C_1}.$$

By transposing the second term on the right to the left of (4·2) the equation can be identified with

$$D_{1,i-1} = h(D_{1,i-1}),$$

so that Newton's iterative formula

$$D_{1,i} = D_{1,i-1} - \frac{h(D_{1,i-1})}{h'(D_{1,i-1})}$$

can be used to solve for D_1. Note also

$$h'(D_1) = \left(1 - \frac{n}{n-1} D_1^2\right)^{\frac{1}{2}(n+\nu-4)}.$$

The initial value of D_1 used to start the iteration procedure was $D_{1,0} = B_1$, for B_1 as given above.

While D_1 is an upper bound for D_α, a lower bound by (4·1) satisfies

$$n\Pr(b_i > D_2) - \binom{n}{2}\Pr(b_i > D_2, b_j > D_2) = \alpha.$$

A first approximation $D_{2,0}$ to D_2 is, therefore, given by

$$n \Pr(b_i > D_{2,0}) = \alpha + \binom{n}{2} \Pr(b_i > D_1, b_j > D_1). \tag{4·3}$$

On replacing D_1 in (4·3) by $D_{2,0}$ a second approximation $D_{2,1}$ is obtained. The process can be continued until $D_{2,t+1}$ and $D_{2,t}$ agree to three decimal places. In the present case $D_{2,0}$ was found to be sufficiently accurate in all but a few cases.

The second term on the right side of equation (4·3) is evaluated by numerical integration. The joint density $f(b_1, b_2)$ given by (3·3) is integrated over the region for which $b_1 > D_1$ and $b_2 > D_1$. This region is shaded in Fig. 1. This numerical integration was performed on an I.B.M. 650 computer. The numerical method used is equivalent to fitting an increasing number of planes to the density surface until the desired accuracy is achieved.

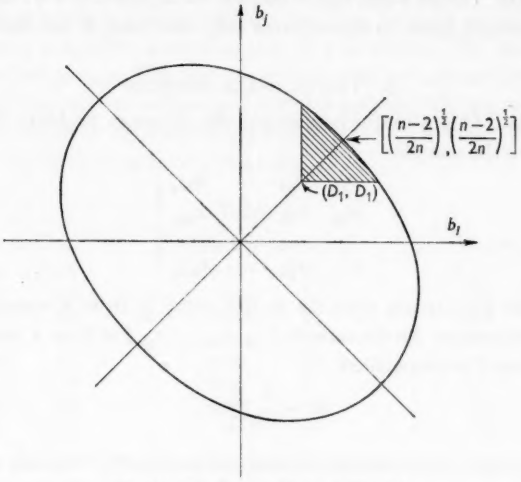

Fig. 1. Positive region of $f(b_i, b_j)$.

An examination of Fig. 1 shows that if $D_1 > [(n-2)/2n]^{\frac{1}{2}}$ then the second term on the right side of (4·3) is zero. Then $d_1 = D_2 = D_\alpha$ is the exact percentage point of b. This is important in that it allows the exact calculation of a number of percentage points for lower values of n and ν.

The lower and upper bounds for the percentage points were found to agree so well for the values of α considered here (0·05, 0·01) that only one value had to be tabulated. Tables 1 and 2 give the 5 and 1 % points, respectively, for selected values of ν and n.

It will be noted that the entries in the row $\nu = 0$, when divided by $\sqrt{n}$, correspond to those given by Chandrasekhar & Pearson (1936, p. 318) for the upper 5 and 1 % points of τ.

(*b*) *The two-sided case, b**

Essentially the same procedure is used to obtain the significance points of b^* as for b. The Bonferroni inequalities in this case give

$$n \Pr(|b_i| > D) - \binom{n}{2} \Pr(|b_i| > D, |b_j| > D) \leqslant \Pr(b^* > D) \leqslant n \Pr(|b_i| > D). \tag{4·4}$$

From the symmetry of $f(b_i)$ we have

$$\Pr(|b_i| > D) = 2\Pr(b_i > D).$$

Let D_α^* be the desired significance point of b^*. Then an upper bound D_1^* can be obtained from (4·2) by replacing α by $\frac{1}{2}\alpha$ in B_1 and solving. A first approximation $D_{2,0}^*$ to a lower bound D_2^* on D_α^* is given by

$$n\Pr(b_i > D_{2,0}^*) = \frac{1}{2}\left[\alpha + \binom{n}{2}\Pr(|b_i| > D_1^*, |b_j| > D_1^*\right]. \tag{4·5}$$

A second approximation $D_{3,0}^*$ can be obtained by replacing D_1^* by $D_{2,0}^*$ in (4·5), etc. The second term on the right of (4·5) is evaluated this time by integrating $f(b_1, b_2)$ over the area in each quadrant where $|b_i| > D_1^*$ and $|b_j| > D_1^*$. The bounds on D_α^* do not agree as well as for the one-sided case. Tables 3 and 4 give bounds for D_α^* for $\alpha = 0{\cdot}05$ and $\alpha = 0{\cdot}01$, respectively. When the bounds agree to three places only one value is tabulated.

5. THE SLIPPAGE PROBLEM

The statistics b and b^* are useful in treating the slippage problem for normal populations. Here we have the sample

$$\left.\begin{array}{cccc} x_{11} & x_{12} & \dots & x_{1n_1} \\ x_{22} & x_{22} & \dots & x_{2n_2} \\ \cdot & \cdot & & \cdot \\ x_{k1} & x_{k2} & \dots & x_{kn_k} \end{array}\right\}. \tag{5·1}$$

We wish to test the hypothesis that the entire array is from a common normal parent against the alternative that the ith sample $(x_{i1}, x_{i2}, \dots, x_{in_i})$ is from a normal parent with a different mean, where i is unspecified.

Let

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$$

and s_l^2 be an independent mean-square estimate of error with t degrees of freedom.

An important special case is that of all equal subsample sizes, i.e. $n_1 = n_2 = \dots = n_k = m$. For this special case the statistics

$$\max_i \frac{\sqrt{m}(\bar{x}_i - \bar{x})}{S_1}, \tag{5·2}$$

and

$$\max_i \frac{\sqrt{m}|\bar{x}_i - \bar{x}|}{S_1}, \tag{5·3}$$

where

$$S_1^2 = \sum_{i=1}^{k}\sum_{j=1}^{m}(x_{ij} - \bar{x})^2 + ts_l^2,$$

are distributed as b and b^*, respectively. The significance points of these statistics are obtained from the tables of b and b^* with the parameters n and ν of the tables replaced by $n = k$ and $\nu = k(m-1)+t$. These slippage tests possess the same desirable properties as do outlier tests based on b and b^* (see Paulson, 1952).

If the same sizes n_i are not all equal but are approximately so the tables of b and b^* can be used to obtain approximate tests. Put

$$\bar{x}_w = \frac{1}{k}\sum_{i=1}^{k}\sqrt{n_i}\,\bar{x}_i, \quad N = \sum_{i=1}^{k} n_i. \tag{5·4}$$

Then the statistics

$$\max_i \frac{\sqrt{n_i}\bar{x}_i - \bar{x}_w}{S_1} \tag{5·5}$$

and

$$\max_i \frac{|\sqrt{n_i}\bar{x}_i - \bar{x}_w|}{S_1} \tag{5·6}$$

give approximate tests based on the tables of b and b^*. The parameters n and ν of the tables are here $n = k$ and $\nu = N + t - k$.

6. Examples

We now give two examples to illustrate the use of the tables.

Example 1

Squibs are small devices for igniting the rocket motors of missiles. Watertightness and shock resistance are important characteristics of squibs. In order to study these characteristics of a large batch a random sample of size 48 was drawn. The sample was randomly subdivided into 3 equal groups. The first group was used as a control unit and received no treatment, the second group was submerged in water and the third group was dropped from a fixed height. Each squib in the entire sample was tested by having a current of 5 amperes passed through it and its time to failure recorded.

Table 6·1

Control (x_{1i})		Watertightness (x_{2i})		Shock (x_{3i})	
0·38	0·51	0·53	0·39	0·51	0·35
·26	·55	·35	·74	·63	·41
·41	·53	·38	·32	·46	·49
·33	·41	·45	·74	·47	·40
·33	·47	1·09	·48	·42	·58
·37	·49	0·46	·37	·45	·46
·54	·42	·57	·52	·41	·38
·76	·34	·47	·44	·39	·48

(Data furnished by Ordnance Missile Laboratories, ARGMA, AOMC, Redstone Arsenal, Alabama.)

$\Sigma x_{1i} = 7{\cdot}10,$	$\Sigma x_{2i} = 8{\cdot}30,$	$\Sigma x_{3i} = 7{\cdot}29,$
$\bar{x}_1 = 0{\cdot}4438,$	$\bar{x}_2 = 0{\cdot}5188,$	$\bar{x}_3 = 0{\cdot}4556,$
$\Sigma x_{1i}^2 = 3{\cdot}3686,$	$\Sigma x_{2i}^2 = 4{\cdot}8768,$	$\Sigma x_{3i}^2 = 3{\cdot}4021,$
$\frac{(\Sigma x_{1i})^2}{16} = 3{\cdot}1506,$	$\frac{(\Sigma x_{2i})^2}{16} = 4{\cdot}3056,$	$\frac{(\Sigma x_{3i})^2}{16} = 3{\cdot}3215,$
$SS(1) = 0{\cdot}2180,$	$SS(2) = 0{\cdot}5712,$	$SS(3) = 0{\cdot}0806.$

From investigation of a large series of similar data it has been found that these delay times are approximately normally distributed, but that for reasons not fully understood, occasionally extremely large delay times occur. Because of this an outliers test was used on each subgroup to pick out such divergent observations. The variance of the bulk of normal observations was assumed to be constant throughout the experiment. The data are given in Table 6·1.

First, we test each of the subgroups for outlying observations, in order to avoid such observations vitiating comparisons among treatments. For each test the variance estimates from the two remaining subgroups are used as an independent estimate of error. For the control group we have from (2·1)

$$b = \frac{0{\cdot}76 - 0{\cdot}4438}{\sqrt{0{\cdot}8698}} = 0{\cdot}3390.$$

Table 1 gives the 5 % point for $n = 16$ and $\nu = 30$ as approximately $b\,(0{\cdot}05;\ 16, 30) = 0{\cdot}384$, so the above value does not attain significance.

For the watertightness group

$$b = \frac{1{\cdot}09 - 0{\cdot}5188}{0{\cdot}93274} = 0{\cdot}6124$$

a value which Table 2 shows to be clearly significant at the 0·01 level. The observation 1·09 is therefore discarded.

Now, since this observation was unduly inflating the variance estimate for the first test, we shall recompute that test with the observation omitted from the variance estimate. The test statistic becomes

$$b = \frac{0{\cdot}76 - 0{\cdot}4438}{\sqrt{0{\cdot}5217}} = 0{\cdot}4378,$$

and is to be compared with $b(0{\cdot}05;\ 16, 29)$.

Table 1 gives $b(0{\cdot}05;\ 15, 24) = 0{\cdot}413$ and increasing either n or ν tends to decrease the percentage point, so the above statistic is significant at the 0·05 level but not at the 0·01 level. The observation 0·76 is omitted from the control group, the sum of squares based on the remaining 15 observations being 0·1113.

For testing the 3rd (shock) group $b = 0{\cdot}2707$. This is not significant at the 0·05 level. Further tests in the subsamples lead to no more discarded observations.

The purpose of the experiment is to test the significance of the water and shock treatments. We are interested in testing the hypothesis that either one or both treatments increased the mean delay time. A two-sided test is appropriate here. For the two-sided test will have a probability of rejection higher than for the null situation under any alternative except when one treatment effect is exactly twice the other (both non-zero).

Since the subsample sizes are large and nearly equal, the approximate test discussed in § 5 should give accurate results. Here $n_1 = n_2 = 15$ and $n_3 = 16$. The weighted means are $\sqrt{n_1}\,\bar{x}_1 = 1{\cdot}637$, $\sqrt{n_2}\,\bar{x}_2 = 1{\cdot}862$, $\sqrt{n_3}\,\bar{x}_3 = 1{\cdot}822$ and $\bar{x}_w = 1{\cdot}774$. Then (5·6) gives

$$\frac{1{\cdot}774 - 1{\cdot}637}{0{\cdot}6442} = 0{\cdot}213.$$

Table 3 gives $b(0{\cdot}05;\ 3, 40) \geqslant 0{\cdot}283$ so the value 0·213 is not significant at the 0·05 level. We conclude that the treatment effects if present, are only small.

Example 2

A sample of six observations was drawn from a table of random normal numbers and a randomly selected observation was increased by two standard deviations. The observations obtained were 265, 223, 291, 105, 43 and 477. A sample of six observations was drawn from a table with the same variance but with a different mean to give an independent estimate of variance. These observations were 171, 111, 185, 271, 68 and 217. The mean and sum of squares about the mean for the first set of observations are 234 and 116,504 respectively. The sum of squares about the mean for the second set is 26,519. So the one-sided test statistic is from (2·1)

$$b = \frac{477 - 234}{\sqrt{143{,}023}} = 0{\cdot}643.$$

Table 1 gives $b(0{\cdot}05; 6, 5) = 0{\cdot}638$, so that the observation 477 is rejected at the 0·05 level. Table 3 gives $b^*(0{\cdot}05; 6, 5) = 0{\cdot}681$, so that the observation 477 is not rejected at the 0·05 level by the two-sided test.*

References

Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*, **2**, 123–47.

Basu, D. (1955). On statistics independent of a complete sufficient statistic. *Sankhyā*, **15**, 377–80.

Borenius, G. (1958). On a distribution of the extreme values in a sample from a normal population. *Skand. Aktuarietidskrift*, **3**, 131–66.

David, H. A. (1956). On the applications to statistics of an elementary theorem in probability. *Biometrika*, **43**, 85–91.

Doornbos, R., Kesten, H. & Prins, H. J. (1956). A class of slippage tests. *Report* S 206 (VP 8), Mathematisch Centrum, Amsterdam.

Grubbs, F. E. (1950). Sample criteria for testing outlying observations. *Ann. Math. Statist.* **21**, 27–58.

Halperin, M., Greenhouse, S., Cornfield, J. & Zalokar, Julia (1955). Tables of percentage points for the Studentized maximum absolute deviate in normal samples. *J. Amer. Statist. Ass.* **50**, 185–95.

Kudô, A. (1956). On the testing of outlying observations. *Sankhyā*, **17**, 67–76.

McKay, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe. *Biometrika*, **27**, 466–72.

Nair, K. R. (1948). The distribution of the extreme deviate from the sample mean and its Studentized form. *Biometrika*, **35**, 118–44.

Paulson, E. (1952). An optimum solution to the k-sample slippage problem for the normal distribution. *Ann. Math. Statist.* **23**, 610–16.

Pearson, E. S. & Chandrasekhar, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. *Biometrika*, **28**, 308–20.

Ruben, H. (1954). On the moments of order statistics in samples from normal populations. *Biometrika*, **41**, 200–27.

Thompson, W. R. (1935). On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *Ann. Math. Statist.* **6**, 214–19.

Appendix

The tables

Tables 1 and 2 give the 5 and 1 % points, respectively, of b [see (2·1)]. The 1 % points are correct to 1 unit in the fourth place. Only a few values for n and ν large are questionable at all in the last digit. The 5 % points of b are correct to three places except for a few large values of n and ν. For $n = 20$ and $\nu = 50$ the value given may be as much as 2 units too large in the third place. No other values in the 5 % table are incorrect by more than one unit in the third place.

Tables 3 and 4 give lower and upper bounds for the percentage points of b^* [see (2·2)]. For each combination of parameter values, lower and upper bounds are given except when these bounds agree to three decimal places and then only one value is given.

When an asterisk (*) replaces the lower bound, the correction term in equation (4·5) is zero, i.e. (upper–lower) = 0. When a horizontal bar (—) replaces the lower bound, the correction term in (4·5) is not zero but so small that the lower agrees with the upper bound to three places, i.e. (upper–lower) $< 0{\cdot}001$.

[* This example was introduced simply to illustrate the method of calculation of the statistic and the use of the tables; presumably when dealing with a real problem it would be clear to the statistician whether a single or two-sided test was the more appropriate. Ed.]

Table 1. *Upper 5% points of* $(x_{max}-\bar{x})/S$

ν \ n	3	4	5	6	7	8	9	10	12	15	20
0	0·8154	0·844	0·836	0·815	0·791	0·768	0·746	0·725	0·689	0·644	0·586
1	0·789	0·800	0·789	0·771	0·752	0·733	0·714	0·697	0·666	0·626	0·574
2	·741	·752	·745	·732	·717	·701	·686	·672	·644	·609	·562
3	·692	·707	·705	·697	·686	·673	·661	·648	·625	·594	·550
4	·648	·668	·671	·666	·658	·648	·638	·627	·607	·579	·540
5	·610	·634	·640	·638	·653	·625	·617	·608	·591	·566	·530
6	0·577	0·604	0·613	0·614	0·610	0·604	0·598	0·591	0·575	0·553	0·520
7	·549	·578	·589	·591	·590	·586	·580	·574	·561	·541	·511
8	·524	·554	·567	·571	·571	·569	·564	·559	·548	·530	·502
9	·502	·533	·547	·553	·554	·553	·550	·546	·536	·520	·494
10	·483	·515	·530	·536	·539	·538	·536	·533	·524	·510	·486
12	0·450	0·482	0·499	0·507	0·511	0·512	0·511	0·509	0·503	0·492	0·472
15	·411	·443	·461	·471	·476	·479	·480	·479	·476	·468	·452
20	·363	·395	·413	·425	·431	·436	·438	·439	·439	·435	·424
24	·335	·366	·384	·396	·403	·408	·411	·413	·415	·413	·405
30	·303	·332	·350	·362	·370	·375	·379	·382	·385	·385	·381
40	·266	·292	·309	·321	·329	·335	·339	·342	·347	·349	·349
50	·239	·264	·280	·291	·299	·305	·310	·313	·318	·322	·323

Table 2. *Upper 1% points of* $(x_{max}-\bar{x})/S$

ν \ n	3	4	5	6	7	8	9	10	12	15	20
0	0·8165	0·8617	0·8739	0·8695	0·8566	0·8394	0·8211	0·8032	0·7687	0·7228	0·6614
1	0·8111	0·8431	0·8478	0·8400	0·8263	0·8104	0·7942	0·7780	0·7465	0·7048	0·6483
2	·7904	·8155	·8176	·8094	·7971	·7833	·7688	·7541	·7260	·6879	·6356
3	·7614	·7844	·7865	·7800	·7698	·7579	·7450	·7320	·7070	·6723	·6239
4	·7299	·7532	·7570	·7527	·7444	·7341	·7229	·7116	·6890	·6576	·6127
5	·6990	·7238	·7297	·7274	·7207	·7120	·7026	·6926	·6724	·6438	·6020
6	0·6703	0·6968	0·7045	0·7037	0·6987	0·6918	0·6837	0·6748	0·6548	0·6306	0·5917
7	·6442	·6720	·6812	·6819	·6786	·6729	·6659	·6581	·6422	·6182	·5822
8	·6204	·6491	·6597	·6620	·6599	·6554	·6495	·6428	·6286	·6066	·5727
9	·5986	·6282	·6401	·6436	·6424	·6389	·6341	·6284	·6156	·5956	·5638
10	·5788	·6091	·6219	·6263	·6263	·6237	·6198	·6148	·6031	·5851	·5556
12	0·5441	0·5723	0·5895	0·5956	0·5971	0·5962	0·5935	0·5899	0·5808	0·5654	0·5398
15	·5017	·5333	·5489	·5566	·5600	·5607	·5597	·5576	·5513	·5393	·5183
20	·4482	·4795	·4962	·5055	·5106	·5132	·5140	·5136	·5104	·5031	·4880
24	·4158	·4463	·4633	·4732	·4792	·4826	·4844	·4850	·4837	·4785	·4668
30	·3779	·4073	·4240	·4346	·4412	·4455	·4480	·4496	·4501	·4479	·4393
40	·3328	·3600	·3763	·3869	·3940	·3990	·4023	·4047	·4071	·4074	·4038
50	·3006	·3162	·3416	·3519	·3591	·3642	·3681	·3711	·3744	·3766	·3751

Note: $S^2 = \sum_1^n (x_i-\bar{x})^2 + \sum_1^{\nu+1} (y_i-\bar{y})^2$, with y_i independent of x_i.

Table 3. *Bounds for upper 5 % points of* $\max|x_i-\bar{x}|/S$

ν \ n	3	4	5	6	7	8	9	10	12	15	20
0	0·8162	0·855	0·857	0·844	0·825	0·804	0·783	0·763	0·727	0·681	0·621
	*	*	*	*	*	*	*	*	*	—	—
1	0·803	0·824	0·820	0·807	0·789	0·771	0·753	0·736	0·704	0·663	0·608
	*	*	*	*	*	*	*	*	—	—	—
2	0·769	0·786	0·782	0·771	0·757	0·741	0·726	0·711	0·683	0·646	0·596
	*	*	*	*	*	*	*	*	—	—	—
3	0·729	0·747	0·746	0·738	0·727	0·714	0·701	0·688	0·664	0·630	0·584
	*	*	*	*	*	*	—	—	—	—	—
4	0·690	0·711	0·713	0·708	0·670	0·689	0·678	0·667	0·646	0·616	0·573
	—	*	*	*	—	—	—	—	—	—	—
5	0·655	0·678	0·684	0·681	0·675	0·667	0·658	0·648	0·629	0·602	0·563
	·654	—	—	—	—	—	—	—	—	—	—
6	0·623	0·649	0·657	0·657	0·652	0·646	0·638	0·630	0·613	0·589	0·553
	·620	—	—	—	—	—	—	—	—	—	—
7	0·596	0·623	0·633	0·635	0·632	0·627	0·621	0·614	0·599	0·577	0·544
	·590	·622	—	—	—	—	—	—	—	—	—
8	0·571	0·600	0·611	0·614	0·613	0·609	0·604	0·598	0·586	0·566	0·535
	·563	·598	·610	—	—	—	—	—	—	—	—
9	0·549	0·579	0·592	0·596	0·596	0·593	0·589	0·584	0·573	0·555	0·526
	·540	·576	·590	·595	·595	—	—	—	—	—	—
10	0·529	0·560	0·573	0·579	0·580	0·579	0·575	0·571	0·561	0·545	0·518
	·519	·556	·572	·578	·579	·578	—	—	—	—	—
12	0·495	0·526	0·542	0·549	0·551	0·551	0·550	0·547	0·539	0·526	0·503
	·483	·522	·539	·547	·550	·550	·549	·546	·539	·525	·502
15	0·454	0·486	0·502	0·511	0·515	0·517	0·517	0·515	0·511	0·501	0·483
	·441	·480	·499	·509	·514	·516	·516	·514	·510	·500	·482
20	0·403	0·434	0·452	0·462	0·468	0·471	0·473	0·473	0·472	0·466	0·453
	·389	·427	·447	·459	·466	·469	·471	·472	·471	·465	·452
24	0·373	0·403	0·421	0·432	0·438	0·442	0·445	0·446	0·446	0·443	0·433
	·359	·395	·416	·428	·436	·440	·443	·444	·445	·442	·432
30	0·338	0·367	0·384	0·395	0·403	0·408	0·411	0·413	0·415	0·414	0·408
	·324	·359	·379	·391	·399	·405	·408	·411	·413	·413	·407
40	0·297	0·323	0·340	0·351	0·359	0·364	0·368	0·371	0·374	0·376	0·374
	·283	·315	·334	·346	·355	·361	·365	·368	·372	·375	·372
50	0·267	0·292	0·308	0·319	0·326	0·332	0·336	0·339	0·344	0·347	0·347
	·254	·284	·302	·314	·323	·328	·333	·336	·341	·345	·345

Note: $S^2 = \sum_1^n (x_i-\bar{x})^2 + \sum_1^{\nu+1} (y_i-\bar{y})^2$, with y_i independent of x_i.

* indicates (upper–lower bound) = 0.
— indicates (upper–lower bound) < 0·001.

Table 4. *Bounds for upper* 1 % *points of* $\max|x_i-\bar{x}|/S$

ν \ n	3	4	5	6	7	8	9	10	12	15	20
0	0·8165	0·864	0·881	0·882	0·874	0·860	0·843	0·827	0·794	0·750	0·688
	*	*	*	*	*	*	*	*	*	*	—
1	0·814	0·851	0·862	0·858	0·847	0·833	0·819	0·804	0·773	0·732	0·674
	*	*	*	*	*	*	*	*	*	*	—
2	0·800	0·830	0·837	0·831	0·821	0·808	0·795	0·781	0·753	0·715	0·662
	*	*	*	*	*	*	*	*	*	*	—
3	0·778	0·805	0·809	0·805	0·796	0·785	0·772	0·759	0·735	0·700	0·651
	*	*	*	*	*	*	*	*	*	—	—
4	0·751	0·777	0·782	0·779	0·772	0·762	0·751	0·740	0·717	0·686	0·639
	*	*	*	*	*	*	*	*	*	—	—
5	0·724	0·750	0·757	0·755	0·749	0·741	0·731	0·721	0·701	0·671	0·628
	*	*	*	*	*	*	*	*	—	—	—
6	0·698	0·725	0·733	0·733	0·728	0·721	0·713	0·704	0·685	0·658	0·617
	—	*	*	*	*	*	*	—	—	—	—
7	0·673	0·702	0·711	0·712	0·708	0·703	0·695	0·687	0·671	0·646	0·608
	—	—	—	*	—	—	—	—	—	—	—
8	0·651	0·680	0·690	0·692	0·690	0·685	0·679	0·672	0·657	0·634	0·599
	—	—	—	—	—	—	—	—	—	—	—
9	0·629	0·659	0·671	0·674	0·673	0·669	0·664	0·657	0·644	0·623	0·591
	—	—	—	—	—	—	—	—	—	—	—
10	0·610	0·640	0·653	0·657	0·657	0·654	0·649	0·644	0·632	0·612	0·581
	·609	—	—	—	—	—	—	—	—	—	—
12	0·576	0·607	0·621	0·626	0·628	0·626	0·623	0·619	0·609	0·593	0·565
	·573	—	—	—	—	—	—	—	—	—	—
15	0·533	0·564	0·580	0·587	0·590	0·590	0·588	0·586	0·579	0·566	0·544
	·529	—	—	—	—	—	—	—	—	—	—
20	0·478	0·509	0·526	0·534	0·539	0·541	0·542	0·541	0·537	0·528	0·512
	·473	·508	·525	—	—	—	—	—	—	—	—
24	0·444	0·475	0·492	0·501	0·507	0·510	0·511	0·511	0·509	0·503	0·490
	·439	·473	·491	—	—	—	—	—	—	—	—
30	0·405	0·434	0·451	0·461	0·468	0·471	0·474	0·475	0·475	0·471	0·462
	·399	·432	·450	·460	·467	·470	·473	·474	·474	—	—
40	0·357	0·385	0·401	0·411	0·418	0·423	0·426	0·428	0·430	0·429	0·424
	·351	·382	·399	·410	·417	·422	·426	·427	·429	—	—
50	0·323	0·349	0·365	0·375	0·382	0·387	0·390	0·393	0·396	0·397	0·395
	·317	·346	·363	·373	·381	·386	·389	·392	·395	·397	·394

Note: $S^2=\sum_1^n(x_i-\bar{x})^2+\sum_1^{\nu+1}(y_i-\bar{y})^2$, with y_i independent of x_i.

* indicates (upper–lower bound) = 0.

— indicates (upper–lower bound) < 0·001.

Biometrika (1961), **48**, 3 *and* 4, *p.* 391
With 2 text-figures
Printed in Great Britain

The ergodic behaviour of random walks

BY J. F. C. KINGMAN
Statistical Laboratory, University of Cambridge

1. Many stochastic processes occurring in practice may be formulated as Markov chains with an enumerable state space. It is then important to know whether or not the chain is ergodic, i.e. whether or not a stationary distribution exists. For particular problems this has often been determined by complex and ingenious methods, a good example being the analysis by Kiefer & Wolfowitz (1955) of the many-server queue. However, Foster (1953) has given a general criterion for a chain to be ergodic, and the purpose of this paper is to examine the way in which his result may be applied to particular processes. By way of example, the technique is applied to two important problems in queueing theory. The results obtained also have consequences in the theory of multi-dimensional random walks, in which context ergodicity simply means that the particle does not escape to infinity.

The processes with which we will be concerned are those which may be formulated as Markov chains on the state space of vectors

$$\mathbf{x} = (x_1, \ldots, x_n),$$

where the x_i are integers. We shall assume that the chain is irreducible and aperiodic, and we shall denote the probability of a transition from $\mathbf{x}$ to $\mathbf{x}+\mathbf{y}$ by $p(\mathbf{x}, \mathbf{y})$. Then we have the following result, due to Foster (1953, Theorems 2 and 3):

The chain is ergodic if and only if there exists a positive function $\phi(\mathbf{x})$ *such that, for all* $\mathbf{x}$

$$\sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y})\, \phi(\mathbf{x}+\mathbf{y}) < \infty \tag{1}$$

and such that, for all but a finite number of $\mathbf{x}$,

$$\sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y})\, \phi(\mathbf{x}+\mathbf{y}) \leqslant \phi(\mathbf{x}) - 1. \tag{2}$$

(Foster states his theorem for the case where (2) holds for all $\mathbf{x} \neq \mathbf{0}$. His proof, however, goes through unchanged in this slightly more general case.)

If we have a particular chain, and we wish to prove it ergodic, it is necessary to construct a function $\phi(\mathbf{x})$ satisfying (1) and (2). In the simple cases considered by Foster, a simple linear function of $\mathbf{x}$ sufficed, but, as will appear later, more complicated functions are often necessary. Our object is to give a method for constructing ϕ. This will be done in a heuristic manner, and it should be emphasized that it is necessary, having constructed ϕ, to verify that it does indeed satisfy Foster's criterion. There are, however, certain special cases in which a more precise analysis can be given.

The technique to be described is most fruitful when the process is a random walk in some region, possibly with complicated boundary conditions. It is, however, by no means restricted to this case.

2. In random walks, a common situation will be that the length of a step is bounded, so that, for any $\mathbf{x}$, $p(\mathbf{x}, \mathbf{y}) = 0$ for all but finitely many $\mathbf{y}$. In this case (1) will be trivially satisfied. To examine (2), consider the effect of approximating $\phi(\mathbf{x}+\mathbf{y})$ by

$$\phi(\mathbf{x}+\mathbf{y}) = \phi(\mathbf{x}) + \mathbf{y}.\operatorname{grad}\phi(\mathbf{x}). \tag{3}$$

Here we are regarding ϕ as a (differentiable) function of the n real variables x_i, and, as usual, grad ϕ is the vector whose ith component is $\partial\phi/\partial x_i$. Then (2) becomes

$$\boldsymbol{\mu}.\operatorname{grad}\phi(\mathbf{x}) \leqslant -1, \tag{4}$$

where

$$\boldsymbol{\mu}(\mathbf{x}) = \sum_{\mathbf{y}} \mathbf{y} p(\mathbf{x}, \mathbf{y}) \tag{5}$$

will be called the *drift* vector at $\mathbf{x}$.

If we now look at the level surfaces of $\phi(\mathbf{x})$, we see that, in order that $\phi(\mathbf{x})$ should always be positive and that $|\operatorname{grad}\phi|$ be bounded away from zero, each level surface must enclose a region, which will usually be bounded, in which ϕ takes smaller values than it does on the surface. In order that $\boldsymbol{\mu}.\operatorname{grad}\phi$ be negative, $\boldsymbol{\mu}(\mathbf{x})$ must point into the region enclosed by the level surface through $\mathbf{x}$.

Hence, given the configuration of drift vectors, we have to construct a family of surfaces in the region of the walk which are nested inside one another, and such that $\boldsymbol{\mu}(\mathbf{x})$ points into the interior of the surface through $\mathbf{x}$. We then choose $\phi(\mathbf{x})$ as a suitable monotonic labelling of the surfaces. This is a purely geometrical problem, and whether or not it is possible depends on the configuration of the drift vectors.

By looking at these vectors, we can gain an idea as to whether or not the chain is likely to be ergodic. For example, if $n = 1$, and the walk is restricted to $x_1 \geqslant 0$, we will expect the chain to be ergodic if, for all $\mathbf{x}$, $\mu_1(\mathbf{x}) < 0$. That this is in fact so, has been shown, in a particular case, by Lindley (1952). This method has, however, relatively little interest in the case $n = 1$, when the strong law of large numbers can be readily invoked.

3. A less trivial illustration is that of a random walk in the first quadrant of the plane. For simplicity we assume that the steps are of unit length, and have the same distributions at all points of the interior $x_1 > 0$, $x_2 > 0$. The boundary conditions are arbitrary except that they are constant along each straight segment of the boundary. (It is straightforward to extend the analysis to a more general situation.)

Under these assumptions we have

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{x}) &= \boldsymbol{\mu} \quad (x_1 > 0, x_2 > 0) \\ &= \boldsymbol{\mu}' \quad (x_1 > 0, x_2 = 0) \\ &= \boldsymbol{\mu}'' \quad (x_1 = 0, x_2 > 0), \end{aligned}$$

where $\boldsymbol{\mu}$, $\boldsymbol{\mu}'$, $\boldsymbol{\mu}''$ are vectors, with

$$\mu_2' \geqslant 0, \quad \mu_1'' \geqslant 0.$$

Now consider the level-curves of $\phi(\mathbf{x})$. These must meet $x_2 = 0$ at such an angle that $\boldsymbol{\mu}$, $\boldsymbol{\mu}'$ both point inside them. This is only possible if the directions of $\boldsymbol{\mu}$, $\boldsymbol{\mu}'$ satisfy a certain condition. A similar consideration on $x_1 = 0$ gives a condition between $\boldsymbol{\mu}$ and $\boldsymbol{\mu}''$. We distinguish three cases.

(i) $\mu_1 < 0$, $\mu_2 < 0$.

In this case the configuration of the level-curves will be as shown in Fig. 1, and the condition for the construction to be possible is clearly

$$\frac{-\mu_1}{-\mu_2} > \frac{\mu_1'}{\mu_2'} \quad \text{and} \quad \frac{-\mu_2}{-\mu_1} > \frac{\mu_2''}{\mu_1''}.$$

The form of the curves suggests that $\phi(\mathbf{x})$ should be a quadratic function of $\mathbf{x}$, and, in fact, it is not difficult to show that

$$\phi(\mathbf{x}) = ax_1^2 - 2bx_1x_2 + cx_2^2$$

satisfies the conditions of Foster's result, so long as a, b, c are chosen so that

$$\frac{\mu_1'}{\mu_2'} < \frac{b}{a} < \frac{-\mu_1}{-\mu_2}, \quad \text{and} \quad \frac{\mu_2''}{\mu_1''} < \frac{b}{c} < \frac{-\mu_2}{-\mu_1}.$$

Hence this simple geometrical argument gives in a rigorous manner, a sufficient condition for the walk to be ergodic. We note that the fact that grad $\phi \to 0$ as $\mathbf{x} \to 0$ does not affect the argument, as equation (2) may fail to hold at $\mathbf{x} = 0$.

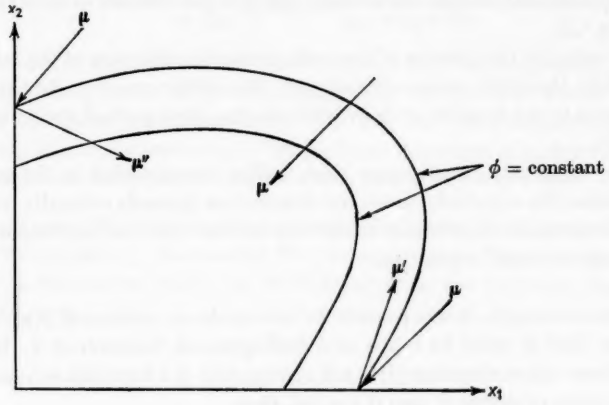

Fig. 1

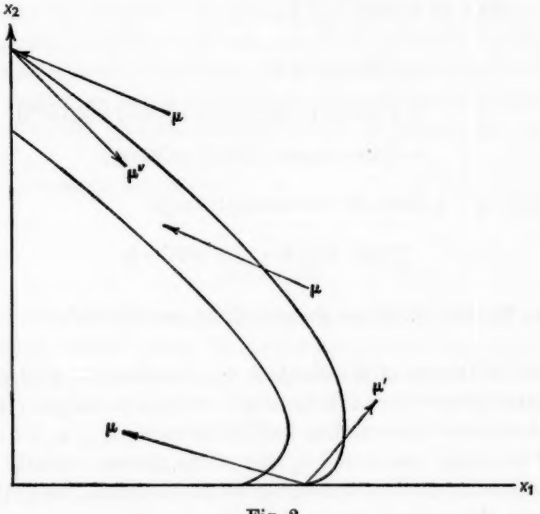

Fig. 2

(ii) $\mu_1 < 0$, $\mu_2 < 0$.

From Fig. 2 the condition is easily seen to be

$$\frac{\mu_2}{-\mu_1} < \frac{-\mu_2''}{\mu_1''}.$$

We see that, in general, a quadratic function will not suffice for $\phi(\mathbf{x})$.

(iii) $\mu_1 < 0$, $\mu_2 > 0$.

In this case it is clear that, however we draw the level-curves, $\boldsymbol{\mu}$ will always point away from their interior.

We may summarize the results in the following statement, which is broadly true for other multi-dimensional random walks (although it is not difficult to construct pathological counter-examples):

If the drift vector in the interior of the walk points to some part of the boundary, and if, when we consider the drift vectors off that part, they either converge (in the sense that any two vectors meet in the interior) or point towards the other part of the boundary, then the walk is ergodic.

This has the important consequence that, unlike the situation in the one-dimensional walk, the criterion for ergodicity in several dimensions depends critically on the boundary conditions. For instance, in the case dealt with above, even the conditions $\mu_1 < 0$, $\mu_2 < 0$ were not enough to ensure ergodicity.

4. In the above example, it was possible to take the level-surfaces of $\phi(\mathbf{x})$ to be geometrically similar, so that ϕ could be taken as a homogeneous function of $\mathbf{x}$. In this case the status of the linear approximation (3) is not merely that of a heuristic convenience. In fact, if ϕ is homogeneous of degree k, and if $x = |\mathbf{x}|$, then

$$\begin{aligned}\phi(\mathbf{x}+\mathbf{y}) &= x^k \phi\left(\frac{\mathbf{x}+\mathbf{y}}{x}\right) \\ &= x^k \phi(\mathbf{x}/x + \mathbf{y}/x) \\ &= x^k [\phi(\mathbf{x}/x) + (\mathbf{y}/x).\operatorname{grad}\phi(\mathbf{x}/x) + O(1/x^2)] \\ &= \phi(\mathbf{x}) + \mathbf{y}.\operatorname{grad}\phi(\mathbf{x})\{1 + O(1/x)\}.\end{aligned}$$

Hence, if $\boldsymbol{\mu}.\operatorname{grad}\phi(\mathbf{x}) \leqslant -1$, then, for x sufficiently large,

$$\sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y})\,\phi(\mathbf{x}+\mathbf{y}) \leqslant \phi(\mathbf{x}) - \tfrac{1}{2}$$

and, replacing ϕ by 2ϕ, the conditions for ergodicity are satisfied.

5. As an example of the use of this method, we consider the problem of two queues in parallel. For a detailed description of this problem, we refer to Haight (1958) and Kingman (1961). It may be formulated as a random walk in the sector $x_1 \geqslant x_2 \geqslant 0$ of the plane, where x_1 is the length of the longer queue and x_2 that of the shorter. Strictly speaking, this is a walk in continuous time, but it is not difficult to see (cf. Kingman, 1961), that, under suitable regularity conditions, the analysis carries through as before.

The transitions of the process are as follows (with mean service time 1, and arrival rate 2ρ)

$$\begin{aligned}
(x_1, x_2) &\to (x_1, x_2+1): \quad \text{intensity } 2\rho \quad (x_1 > x_2),\\
(x_1, x_2) &\to (x_1+1, x_2): \quad \text{intensity } 2\rho \quad (x_1 = x_2),\\
(x_1, x_2) &\to (x_1-1, x_2): \quad \text{intensity } 1 \quad (x_1 > x_2),\\
(x_1, x_2) &\to (x_1, x_2-1): \begin{cases} \text{intensity } 1 & (x_1 > x_2 > 0),\\ \text{intensity } 2 & (x_1 = x_2 > 0). \end{cases}
\end{aligned}$$

Hence we have

$$\begin{aligned}
\boldsymbol{\mu}(\mathbf{x}) &= (-1, 2\rho-1) \quad (x_1 > x_2 > 0),\\
&= (2\rho, -2) \quad (x_1 = x_2 > 0),\\
&= (-1, 2\rho) \quad (x_1 > x_2 = 0),\\
&= (2\rho, 0) \quad (x_1 = x_2 = 0).
\end{aligned}$$

If $\rho < 1$, it is easy to see that suitable level-curves are $x_1^2 + x_2^2 = \text{constant}$, and in fact it is readily verified that $\phi(\mathbf{x}) = A(x_1^2 + x_2^2)$ satisfies the conditions of Foster's theorem for some A. Hence the queue is ergodic when $\rho < 1$. If it had been attempted to prove this result by guessing ϕ, and, for instance, by trying linear functions of $\mathbf{x}$, this result would not have been proved, and the geometrical standpoint shows why this is so.

6. As a second example, we consider the Kiefer–Wolfowitz (1955) analysis of the many-server queue. For simplicity, we consider the two-server case. There it is shown that the waiting time in a two-server queue can be deduced from the following random walk in the sector $x_1 \geqslant x_2 \geqslant 0$:

$$\mathbf{x} \to \mathbf{x}',$$

where

$$\begin{aligned}
x_1' &= [\max(x_1, x_2+s) - t]^+,\\
x_2' &= [\min(x_1, x_2+s) - t]^+.
\end{aligned}$$

Here s, t represent the service and inter-arrival times, and $y^+ = \max(y, 0)$.

In fact, the x_i are not constrained to take only integer values, but Kiefer & Wolfowitz showed that it was sufficient to consider the case in which x_i, s, t take on values which are the multiples of some number ϖ, and it is clear that, without loss of generality, we may take $\varpi = 1$. The result obtained by these authors is that, so long as $Es < 2Et$ the process is ergodic.

The drift vectors are easily seen to be:

in the interior: $\quad (-\bar{t}, \bar{s}-\bar{t})$,

on $x_1 = x_2$: $\quad (\bar{s}-\bar{t}, -\bar{t})$,

where $\bar{y} = Ey$. At points in the interior near the boundaries, $\boldsymbol{\mu}$ will lie between the value given and the value on the boundary. It is not immediately clear what we should take as the value of $\boldsymbol{\mu}$ on $x_2 = 0$, except that $\mu_1 < 0$, $\mu_2 > 0$. If we had

$$\frac{\mu_2}{-\mu_1} < 1$$

we could clearly use the linear function $\phi(\mathbf{x}) = x_1 + x_2$. In fact, the Kiefer–Wolfowitz analysis depends critically on this quantity, and the boundary condition at $x_2 = 0$ causes

a great deal of difficulty. However, from the geometrical picture, we can see that, whatever the value of $\mu_2-\mu_1$, the function $\phi(\mathbf{x}) = x_1^2+x_2^2$ is suitable. In fact, for fixed x_1, x_2,

$$\begin{aligned} E(x_1^2+x_2^2) &= E[\{(x_1-t)^+\}^2+\{(x_2+s-t)^+\}^2] \\ &\leqslant E[(x_1-t)^2+(x_2+s-t)^2] \\ &= x_1^2+x_2^2-2x_1\bar{t}+2x_2(\bar{s}-\bar{t})+O(1) \quad \text{as} \quad x\to\infty \\ &\leqslant x_1^2+x_2^2-2x_2(2\bar{t}-\bar{s})+O(1). \end{aligned}$$

Hence, taking $\phi(\mathbf{x}) = A(x_1^2+x_2^2)$, for some A, satisfies the conditions of Foster's theorem. It is easy to show that (1) is also satisfied.

For the k-server queue, we use similarly $\phi = A\sum_1^k x_i^2$.

7. We therefore see that Foster's criterion can be a powerful method of proving Markov processes ergodic. Whilst the technique presented in this paper is not in itself rigorous, it nevertheless suggests the methods by which a rigorous proof may be most easily constructed. It breaks down, however, when the mean vector vanishes, as in symmetric walks, and a more delicate treatment is necessary. We might, for instance, use the second-order approximation

$$\phi(\mathbf{x}+\mathbf{y}) = \phi(\mathbf{x})+\Sigma y_j\frac{\partial\phi}{\partial x_j}+\tfrac{1}{2}\Sigma y_i y_j\frac{\partial^2\phi}{\partial x_i\partial x_j}$$

leading to the condition $$\Sigma\mu_{ij}\frac{\partial^2\phi}{\partial x_i\partial x_j}\leqslant -1,$$

where $$\mu_{ij} = \sum_{\mathbf{y}} p(\mathbf{x},\mathbf{y})\,y_i y_j.$$

Another problem to which the present work is not applicable is that of absorbing barriers. In this case the chain is not irreducible, and cannot be ergodic. However, if we can replace the absorbing barriers by others which are such that the chain is irreducible and ergodic then the probability of ultimate absorption is unity, since in an ergodic chain there is probability one of ultimate transition from any one state to any other. The converse of this result is false.

There is, however, a wide class of processes occurring in practice which may be formulated as random walks, and for these it is suggested that the analysis of this paper provides a fruitful method for examining their ergodic properties.

I am indebted to Prof. P. Whittle for his encouragement, and to the Department of Scientific and Industrial Research for a Research Studentship.

REFERENCES

FOSTER, F. G. (1953). On the stochastic matrices associated with certain queueing processes. *Ann. Math. Statist.* **24**, 355–60.

HAIGHT, F. A. (1958). Two queues in parallel. *Biometrika*, **45**, 401–10.

KIEFER, J. & WOLFOWITZ, J. (1955). On the theory of queues with many servers. *Trans. Amer. Math. Soc.* **78**, 1–18.

KINGMAN, J. F. C. (1961). Two similar queues in parallel. *Ann. Math. Statist.* (To appear.)

LINDLEY, D. V. (1952). Theory of queues with a single server. *Proc. Camb. Phil. Soc.* **48**, 277–89.

Biometrika (1961), **48**, 3 *and* 4, *p.* 397
Printed in Great Britain

The goodness-of-fit of a single (non-isotropic) hypothetical principal component

BY A. M. KSHIRSAGAR
*University College London**

1. INTRODUCTION

The goodness-of-fit of a single hypothetical discriminant function in the problem of discrimination between several groups has been considered by Bartlett (1951*a*). Virtually, this is the same as testing the adequacy of a single canonical variable to bring out completely the relationship between two vector variables. A single hypothetical canonical variable will be adequate in this 'canonical' or 'external' analysis, if (i) there is only one non-zero canonical correlation, and (ii) the direction of the hypothetical canonical variable coincides with that of the true canonical variable corresponding to the non-zero canonical correlation. Bartlett (1951*a*) and Williams (1952, 1955) have considered these two aspects of 'collinearity' and 'direction' in detail, and have derived certain exact tests for them.

This situation in 'external' analysis has a parallel in 'internal' or 'principal components' analysis. If the population variance-covariance matrix of a number of variables has all its latent roots equal, except one, then the 'principal component' or latent vector corresponding to this 'anomalous' root can be called the single 'non-isotropic' component because if this component is removed, the variation in any other orthogonal direction is isotropic. In geometrical language, all the principal axes of the familiar ellipsoid (given by the exponent in the density function of a multivariate normal distribution), except one, are equal; in other words, the ellipsoid does not degenerate into a sphere because of this exceptional principal axis. Such a situation can arise in factor-analysis if there is only one 'common' factor (for details see § 2) besides the 'specific' factor in the factor-structure. If, therefore, we are testing the adequacy of a single hypothetical non-isotropic principal component, it is necessary—as in the case of 'external' analysis—to consider the following two aspects: (i) departure from the hypothesis due to there being more than one non-isotropic principal components, and (ii) departure due to deviation in direction of the true principal component from the hypothetical one.

In § 2 of this paper, these two aspects are considered and an over-all χ^2 test is obtained, and the underlying assumptions are stated clearly. In § 3, the over-all χ^2 is partitioned and its 'direction' part χ_d^2 is separated, using a geometrical argument similar to the one used by Bartlett (1951*a*) in deriving the 'direction factor' in external analysis. In § 4, χ_d^2 is expressed in terms of rectangular co-ordinates. In § 5, χ_d^2 is expressed in terms of the sample latent roots and principal components in a manner similar to that of Williams (1955). This is also alternatively expressed in terms of 'residual' roots of the sample variance-covariance matrix, following Williams's (1952) idea of 'residual roots'. In § 6, the problem of the distribution of these residual roots is considered; and finally in § 7, the theory is illustrated by a numerical example and a method of deriving approximate confidence intervals for the elements of the single non-isotropic component is described and illustrated.

* Present address: University of Bombay.

2. Adequacy of a single non-isotropic principal component

Let $\mathbf{x}' = [x_1, x_2, \ldots, x_p]$ be a (row) vector having a p-variate normal distribution with zero means and variance-covariance matrix $\mathbf{\Sigma}$. There is an orthogonal matrix

$$\mathbf{L} = [l_{ij}] = [\mathbf{l}_{(1)} | \mathbf{l}_{(2)} | \ldots | \mathbf{l}_{(p)}]' \tag{2·1}$$

such that

$$\mathbf{\Sigma} = \mathbf{L}' \text{ diag. } (\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)\, \mathbf{L} \tag{2·2}$$

where $\sigma_i^2 (i = 1, 2, \ldots, p)$ are the latent roots of $\mathbf{\Sigma}$ and $\mathbf{l}_{(i)}$ are the corresponding (column) latent vectors; diag. stands for a diagonal matrix, the elements of which are written in the adjoining bracket. If the roots are arranged in descending order of magnitude as

$$\sigma_1^2 \geqslant \sigma_2^2 \geqslant \ldots \geqslant \sigma_p^2,$$

then $\mathbf{l}'_{(i)}\mathbf{x}$ is called the ith principal component. The transformation

$$\mathbf{y}' = [y_1, y_2, \ldots, y_p] = \mathbf{x}'\mathbf{L}' \tag{2·3}$$

to the principal components, shows that the y_i are normal, independent variables with zero means and variances $\sigma_i^2 (i = 1, 2, \ldots, p)$. If $\mathbf{\Sigma}$ has only one latent root σ_1^2 different from all the remaining $(p-1)$ roots which are all equal to $\sigma^2 (< \sigma_1^2)$, say then

$$\begin{aligned} \mathbf{\Sigma} &= \mathbf{L}' \text{ diag. } (\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)\, \mathbf{L} \\ &= (\sigma_1^2 - \sigma^2)\, \mathbf{l}_{(1)} \mathbf{l}'_{(1)} + \sigma^2 \mathbf{I}, \end{aligned} \tag{2·4}$$

where $\mathbf{I}$ is the identity matrix. Thus $\mathbf{l}'_{(1)}\mathbf{x}$ is the only non-isotropic principal component and $\mathbf{\Sigma}$ is completely determined by $\mathbf{l}_{(1)}$, σ_1^2 and σ^2. Lawley (1953) has made use of such a representation of $\mathbf{\Sigma}$ in factor analysis when there is only one 'common' factor in addition to the 'specific' factors, so that the factor-structure is

$$x_i = g_i f + e_i \quad (i = 1, 2, \ldots, p), \tag{2·5}$$

where x_i is the ith test variate, g_i is the 'loading' of the ith test in the common factor f and e_i is the factor specific to the ith test. Lawley assumes that f and e_i are all normal independent variables; the variance of f is unity and the units of measurement are so chosen that the e_i have the same variance σ^2, which is also assumed to be known. Under these assumptions, the population variance-covariance matrix is

$$\mathbf{\Sigma} = \mathbf{g}\mathbf{g}' + \sigma^2 \mathbf{I}, \tag{2·6}$$

where $\mathbf{g}' = [g_1, \ldots, g_p]$ is the vector of loadings. Lawley then proceeds to use (2·4) to obtain a unique definition of $\mathbf{g}$ by defining it as

$$\mathbf{g} = (\sigma_1^2 - \sigma^2)^{\frac{1}{2}}\, \mathbf{l}_{(1)}.$$

Since σ^2 is assumed to be known, there is no loss of generality in taking it to be unity. The modifications in the tests of significance derived in this paper, if σ^2 has some other value, are obvious. If σ^2 is not known, it can be estimated by the average of the $(p-1)$ roots remaining after removing the largest root from the sample estimate of $\mathbf{\Sigma}$ (see Lawley, 1956). Let

$$\underset{n \times p}{\mathbf{X}} = [x_{it}] \quad (i = 1, \ldots, p;\ t = 1, \ldots, n) \tag{2·7}$$

be a sample of size n from the distribution of $\mathbf{x}$. The maximum likelihood estimate of $\mathbf{\Sigma}$ based on n degrees of freedom is $\mathbf{A}/n$, where

$$\mathbf{A} = [a_{ij}] = \mathbf{X}'\mathbf{X} \tag{2·8}$$

is the matrix of the sum of squares and products of the sample values. There is an orthogonal matrix

$$\mathbf{G} = [g_{ij}] = [\mathbf{g}_{(1)}|\mathbf{g}_{(2)}|\ldots|\mathbf{g}_{(p)}]' \tag{2·9}$$

such that

$$\mathbf{A} = \mathbf{G}' \operatorname{diag}.(\theta_1^2, \ldots, \theta_p^2)\,\mathbf{G}, \tag{2·10}$$

where $\theta_1^2 \geqslant \theta_2^2 \geqslant \ldots \geqslant \theta_p^2$ are the latent roots of $\mathbf{A}$, with $\mathbf{g}_{(1)}, \mathbf{g}_{(2)}, \ldots, \mathbf{g}_{(p)}$ as the corresponding latent vectors.

$$\mathbf{z}' = [z_1, z_2, \ldots, z_p] = \mathbf{x}'\mathbf{G}' \tag{2·11}$$

is the transformation from $\mathbf{x}$ to the sample principal components $z_i\,(i = 1, \ldots, p)$.

The assumption that all the remaining roots of $\boldsymbol{\Sigma}$, when the largest root σ_1^2 is removed, are equal can, if desired, be tested by an approximate χ^2 test due to Bartlett (1950, 1951*b*, 1951*c*, 1954). Lawley (1956) later investigated this statistic in considerable detail. Throughout this paper it is assumed that this test gives a non-significant result, thus providing no evidence against the adequacy of a single non-isotropic principal component

$$\mathbf{l}_{(1)}'\mathbf{x}. \tag{2·12}$$

Let the null-hypothesis, now, specify that a hypothetical function $\mathbf{l}'\mathbf{x}$, where

$$\mathbf{l}' = [l_1, \ldots, l_p], \quad \mathbf{l}'\mathbf{l} = 1 \tag{2·13}$$

is the single non-isotropic principal component. If this is true, $\mathbf{l} = \mathbf{l}_{(1)}$, the true principal component. If, therefore, we take any orthogonal matrix $\mathbf{L}$ with $\mathbf{l}'$, as its first row and make the transformation (2·3), then y_1 is normally distributed with zero mean and variance σ_1^2 while $y_i\,(i = 2, \ldots, p)$ are all normal independent variables with zero means and variances $\sigma^2 = 1$; all the y_i are independently distributed.

Hence if

$$\underset{n \times p}{\mathbf{Y}} = [y_{it}] = \mathbf{XL}', \tag{2·14}$$

then

$$\chi^2 = \sum_{i=2}^{p} \sum_{t=1}^{n} y_{it}^2 \tag{2·15}$$

is a χ^2 with $n(p-1)$ degrees of freedom. Let

$$\mathbf{B} = [b_{ij}] = \mathbf{Y}'\mathbf{Y} = \mathbf{LX}'\mathbf{XL}' = \mathbf{LAL}'. \tag{2·16}$$

The above χ^2 in (2·15) can now be expressed as

$$\begin{aligned} \sum_{i=1}^{p} \sum_{t=1}^{n} y_{it}^2 - \sum_{t=1}^{n} y_{1t}^2 &= tr\mathbf{B} - \mathbf{l}'\mathbf{Al} \\ &= tr\mathbf{A} - \mathbf{l}'\mathbf{Al} \end{aligned}$$

as $\mathbf{L}$ is orthogonal. Hence the statistic for testing the over-all goodness-of-fit of the single hypothetical (non-isotropic) principal component $\mathbf{l}'\mathbf{x}$ is

$$\chi^2 = tr\mathbf{A} - \lambda^2, \tag{2·17}$$

where $\lambda^2/n = \mathbf{l}'\mathbf{Al}/n = \sum_{t=1}^{n} y_{1t}^2/n$ is the sample variance of $\mathbf{l}'\mathbf{x}$. Obviously λ^2/σ_1^2 has a χ^2 distribution with n degrees of freedom. This χ^2 of (2·17) measures the over-all departure from the null hypothesis. As already explained in § 1, this departure can be due to (i) there being more than one true non-isotropic principal components, and/or (ii) the hypothetical

component $\mathbf{l}$ being not the same as the true component $\mathbf{l}_{(1)}$, i.e. due to difference in the directions of $\mathbf{l}$ and $\mathbf{l}_{(1)}$. If we are more interested in the direction aspect of the hypothetical component, it is necessary to isolate the contribution χ_d^2 to the total χ^2, due to direction and test it. This is done in the next section.

3. Partitioning of the over-all χ^2

In canonical analysis Bartlett (1951*a*) used an ingenious geometrical method for deriving the 'direction factor'. A similar method, with appropriate changes corresponding to the 'internal' analysis is employed here. If the null-hypothesis is true, i.e. $\mathbf{l} = \mathbf{l}_{(1)}$, then any variate $y_i = \mathbf{l}'_{(i)}\mathbf{x}$, where $\mathbf{l}_{(i)}$ is orthogonal to $\mathbf{l}$ (and $\mathbf{l}'_{(i)}\mathbf{l}_{(i)} = 1$), has unit variance and is independently distributed of $y_1 = \mathbf{l}'\mathbf{x}$. It is easy to see that the *sample projection* of one normal variable on another independent normal variable is also a normal variable. Hence the sample projection of $\mathbf{l}'_{(i)}\mathbf{x}$ on $\mathbf{l}'\mathbf{x}$, viz.

$$\mathbf{l}'_{(i)}\mathbf{A}\mathbf{l}/(\mathbf{l}'\mathbf{A}\mathbf{l})^{\frac{1}{2}},$$

is a standard normal variable. This is true for every $\mathbf{l}_{(i)}$ orthogonal to $\mathbf{l}$. Hence

$$\sum_{i=2}^{p} \frac{(\mathbf{l}'_{(i)}\mathbf{A}\mathbf{l})^2}{\mathbf{l}'\mathbf{A}\mathbf{l}}$$

has a χ^2 distribution with $(p-1)$ degrees of freedom. This is the required direction component χ_d^2 of the total χ^2. This can be better expressed as

$$\begin{aligned}\chi_d^2 &= \sum_{i=1}^{p} \frac{(\mathbf{l}'\mathbf{A}\mathbf{l}_{(i)})\,(\mathbf{l}'_{(i)}\mathbf{A}\mathbf{l})}{\mathbf{l}'\mathbf{A}\mathbf{l}} - \mathbf{l}'\mathbf{A}\mathbf{l} \\ &= \frac{\mathbf{l}'\mathbf{A}\left(\sum_{i=1}^{p} \mathbf{l}_{(i)}\mathbf{l}'_{(i)}\right)\mathbf{A}\mathbf{l}}{\lambda^2} - \lambda^2 \\ &= \frac{\mathbf{l}'\mathbf{A}^2\mathbf{l}}{\lambda^2} - \lambda^2, \end{aligned} \tag{3·1}$$

as $\sum_{i=1}^{p} \mathbf{l}_{(i)}\mathbf{l}'_{(i)} = \mathbf{I}$, the $\mathbf{l}_{(i)}$ being unit orthogonal vectors.

4. χ_d^2 in terms of rectangular co-ordinates

There is also an alternative analytical method of obtaining χ_d^2. The matrix B of (2·16) has the Wishart distribution

$$\text{const.}\,|B|^{\frac{1}{2}(n-p-1)} \exp\{-\tfrac{1}{2}(b_{22}+b_{33}+\ldots+b_{pp}+b_{11}/\sigma_1^2)\}\,d\mathbf{B}, \tag{4·1}$$

where $d\mathbf{B}$ stands for the product of the differentials of the $p(p+1)/2$ distinct elements of $\mathbf{B}$. Make the transformation

$$\mathbf{B} = \mathbf{T}\mathbf{T}', \tag{4·2}$$

where

$$\mathbf{T} = [t_{ij}] \tag{4·3}$$

is a lower triangular matrix (i.e. $t_{ij} = 0$ if $j > i$); the t_{ij} are called rectangular co-ordinates. The Jacobian of transformation (see Deemer & Olkin, 1951) is

$$2^p \prod_{i=1}^{p} t_{ii}^{p+1-i}$$

and the distribution of T is, therefore, (see also Mauldon, 1955)

$$\text{const.} \prod_{i=1}^{p} t_{ii}^{n-i} \exp\left\{-\frac{1}{2\sigma_1^2} t_{11}^2 - \tfrac{1}{2}(tr\mathbf{TT}' - t_{11}^2)\right\} d\mathbf{T}. \tag{4·4}$$

Hence t_{11}^2 is a $\chi^2\sigma_1^2$ with n degrees of freedom, while t_{ii}^2 $(i=2,\ldots,p)$ is a χ^2 with $n-i+1$ degrees of freedom and t_{ij} $(i>j)$ are standard normal variables; all of them are independently distributed. The over-all χ^2 of (2·17) is thus alternatively,

$$t_{21}^2 + t_{22}^2 + \ldots + t_{pp}^2 = tr\mathbf{TT}' - t_{11}^2. \tag{4·5}$$

The required direction component χ_d^2 measuring the departure of the hypothetical function $y_1 = \mathbf{l}'\mathbf{x}$ from the true one, viz. $\mathbf{l}'_{(1)}\mathbf{x}$, due to direction, is

$$\chi_d^2 = t_{21}^2 + t_{31}^2 + \ldots + t_{p1}^2 \tag{4·6}$$

which is a χ^2 with $(p-1)$ degrees of freedom if $\mathbf{l} \equiv \mathbf{l}_{(1)}$. The identity of this expression with (3·1) can be established by equating the elements in the first row of

$$\mathbf{TT}' = \mathbf{LAL}' \tag{4·7}$$

(each is equal to B; see (4·2) and (2·17)). We get

$$\left.\begin{aligned} t_{11}^2 &= \mathbf{l}'\mathbf{Al} = \lambda^2 \\ t_{11}t_{21} &= \mathbf{l}'_{(2)}\mathbf{Al} \\ &\cdots\cdots\cdots\cdots \\ &\cdots\cdots\cdots\cdots \\ t_{11}t_{p1} &= \mathbf{l}'_{(p)}\mathbf{Al} \end{aligned}\right\}. \tag{4·8}$$

Squaring and adding,

$$t_{11}^2(t_{11}^2 + t_{21}^2 + \ldots + t_{p1}^2) = \mathbf{l}'\mathbf{A}\left(\sum_{i=1}^{p} \mathbf{l}_{(i)}\mathbf{l}'_{(i)}\right)\mathbf{Al}$$

$$= \mathbf{l}'\mathbf{A}^2\mathbf{l}$$

as $\mathbf{l}_{(i)}$ are unit orthogonal vectors. Hence

$$t_{21}^2 + t_{31}^2 + \ldots + t_{p1}^2 = \frac{\mathbf{l}'\mathbf{A}^2\mathbf{l}}{\lambda^2} - \lambda^2$$

$$= \chi_d^2. \tag{4·9}$$

5. χ_d^2 IN TERMS OF SAMPLE LATENT ROOTS AND 'RESIDUAL' ROOTS

From (2·3) and (2·11), the relation between the true and sample principal components is

$$\mathbf{y} = \mathbf{LG}'\mathbf{z} = \mathbf{Wz}, \tag{5·1}$$

where

$$\mathbf{LG}' = \mathbf{W} = [w_{ij}]. \tag{5·2}$$

Hence

$$y_1 = \mathbf{l}'_{(1)}\mathbf{x} = w_{11}z_1 + \ldots + w_{1p}z_p \tag{5·3}$$

which, for convenience, will be written simply as $\sum_{i=1}^{p} w_i z_i$ hereafter. It should be noted that

$$\mathbf{WW}' = \mathbf{LG}'\mathbf{GL}' = \mathbf{LL}' = \mathbf{I} \tag{5·4}$$

and from (2·10), (2·14) and (2·16),

$$\mathbf{B} = \mathbf{W}\,\text{diag.}\,(\theta_1^2, \ldots, \theta_p^2)\,\mathbf{W}'. \tag{5·5}$$

From (5·3),
$$\lambda^2 = \mathbf{l}'_{(1)}\mathbf{A}\mathbf{l}_{(1)} = \sum_{i=1}^{p} w_i^2\theta_i^2. \tag{5·6}$$

From (2·10) and (5·2)
$$\mathbf{l}'_{(1)}\mathbf{A}^2\mathbf{l}_{(1)} = \sum_{i=1}^{p} w_i^2\theta_i^4. \tag{5·7}$$

Hence from (3·1)
$$\chi_d^2 = \frac{\sum_{i=1}^{p} w_i^2\theta_i^4}{\sum_{i=1}^{p} w_i^2\theta_i^2} - \sum_{i=1}^{p} w_i^2\theta_i^2. \tag{5·8}$$

Thus, if the hypothetical principal component is given in terms of the sample principal components, as (5·3) and if it is the true one, (5·8) gives a suitable expression for χ_d^2 in terms of the sample latent roots θ_i^2. Bartlett (1951*a*) has derived a similar expression in 'canonical' analysis. This expression is suitable for expressing χ_d^2 in terms of θ_i^2 and $\phi_j^2\,(j=1,2,\ldots,p-1)$, the sample 'residual' roots when the hypothetical principal component $\sum_{i=1}^{p} w_i z_i$ is eliminated This idea of residual roots is due to Williams (1952) and was employed by him in discriminant analysis. $\theta_i\,(i=1,2,\ldots,_p)$ are the principal semi-axes of the ellipsoid

$$z_1^2/\theta_1^2 + \ldots + z_p^2/\theta_p^2 = 1.$$

The section by a $(p-1)$-space normal to the direction $(w_1,\ldots,w_p)$ specified by the hypothetical principal component $\sum_{1}^{p} w_i z_i$ is an ellipsoid with principal semi-axes $\phi_j\,(j=1,\ldots,p-1)$. Williams (1952) has proved that these ϕ's—the residual roots—satisfy the equation

$$\sum_{i=1}^{p} \frac{w_i^2\theta_i^2}{\phi^2-\theta_i^2} = 0 \tag{5·9}$$

of degree $(p-1)$ in ϕ^2. Collecting the coefficients of $(\phi^2)^{p-1}$, $(\phi^2)^{p-2}$, and the constant term, we find
$$\prod_{j=1}^{p-1}\phi_j^2 = \frac{1}{\lambda^2}\sum_{i=1}^{p}\theta_i^2 \tag{5·10}$$

and
$$\sum_{i=1}^{p}\theta_i^2 - \sum_{j=1}^{p-1}\phi_j^2 = \sum_{i=1}^{p} w_i^2\theta_i^4/\lambda^2.$$

Hence from (5·8)
$$\chi_d^2 = \sum_{i=1}^{p}\theta_i^2 - \sum_{j=1}^{p-1}\phi_j^2 - \lambda^2. \tag{5·11}$$

This is the required expression for χ_d^2 in terms of the original roots θ_i^2, the residual roots ϕ_j^2 and λ^2/n, the sample variance accounted for by the hypothetical component. Williams has also proved that
$$w_i^2 = \frac{\lambda^2\prod_k(\phi_k^2-\theta_i^2)}{\theta_i^2\prod_{i\neq j}(\theta_j^2-\theta_i^2)}. \tag{5·12}$$

6. Conditional distribution of the original and the residual roots

In canonical analysis, Williams (1952) has explicitly derived the conditional joint distribution of the original and the residual roots, making use of the sufficiency of the multiple correlation of the single true discriminant function. This property of sufficiency was first noted by Bartlett (1947, 1951*a*). In a recent publication, the present author (1960) has supplied

some missing steps and arguments in Williams's derivation of the distribution. In this section, a similar argument is employed for considering the distribution of the original roots θ_i^2 and the residual roots ϕ_j^2 when λ^2 is fixed.

The distribution of **B** is given by (4·1). Transform from **B** to $\theta_i^2 (i = 1, \ldots, p)$ and **W**, using (5·4) and (5·5), viz.

$$\mathbf{B} = \mathbf{W} \operatorname{diag}.(\theta_1^2, \ldots, \theta_p^2)\, \mathbf{W}',$$

$$\mathbf{W}\mathbf{W}' = \mathbf{I}.$$

The transformation is unique if $w_i \geqslant 0\ (i = 1, \ldots, p)$. Since **W** is orthogonal, only $p(p-1)/2$ elements of **W** are independent. We can take these to be $w_1, w_2, \ldots, w_{p-1}$ and suitable others. (w_p cannot be taken as $\sum_1^p w_i^2 = 1$.) These are called the co-ordinates of **W**. When σ_1^2 in the distribution of **B** is unity, it has been proved (see Anderson, 1958) that the θ_i^2 and **W** are independently distributed. The distribution of θ_i^2 in this case is

$$f_1(\theta_1^2, \ldots, \theta_p^2) \prod_{i=1}^p d\theta_i^2 = \text{const.} \prod_{i=1}^p (\theta_i^2)^{\frac{1}{2}(n-p-1)} \exp\left(-\tfrac{1}{2}\sum_1^p \theta_i^2\right) \prod_{i<j} (\theta_i^2 - \theta_j^2) \prod_{i=1}^p d\theta_i^2. \tag{6·1}$$

Unfortunately, an explicit expression for the distribution of **W** in terms of the co-ordinates of **W** is not available. Halmos (1950) (see also Anderson, 1958, pp. 318–23) has proved that it is the 'Haar Invariant' distribution. Symbolically we may write the distribution as

$$f_2(\mathbf{W})\, d(\mathbf{W}). \tag{6·2}$$

However, when $\sigma_1^2 \neq 1$, the exponent in the distribution of **B** is

$$\begin{aligned} \frac{b_{11}}{\sigma_1^2} + b_{22} + \ldots + b_{pp} &= tr\mathbf{B} + \left(\frac{1}{\sigma_1^2} - 1\right) b_{11} \\ &= \sum_1^p \theta_i^2 + \left(\frac{1}{\sigma_1^2} - 1\right) \sum_1^p w_i^2 \theta_i^2 \\ &= \sum_1^p \theta_i^2 + \left(\frac{1}{\sigma_1^2} - 1\right) \lambda^2 \end{aligned} \tag{6·3}$$

and hence θ_i^2 and **W** are no longer independent, but their joint distribution is

$$\frac{1}{\sigma_1^n} f_1(\theta_1^2, \ldots, \theta_p^2) f_2(\mathbf{W}) \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_1^2} - 1\right)\lambda^2\right\} \prod_1^p d\theta_i^2\, d\mathbf{W}. \tag{6·4}$$

It was noted in § 2, that λ^2/σ_1^2 is a χ^2 with n degrees of freedom and hence the conditional distribution of θ_i^2 and **W** when λ^2 is fixed is

$$\frac{\text{const.}\ f_1(\theta_1^2, \ldots, \theta_p^2) f_2(\mathbf{W})\, d\mathbf{W} \prod_{i=1}^p d\theta_i^2}{e^{-\frac{1}{2}\lambda^2} (\lambda^2)^{\frac{1}{2}n-1} d\lambda^2}. \tag{6·5}$$

Rigorously one of the quantities θ_i^2 and **W** must be expressed in terms of λ^2 using (5·10) but this is not carried out explicitly to preserve symmetry. The distribution (6·5) does not involve σ_1^2, showing that λ^2 is a sufficient statistic. Following Williams's (1952) idea in canonical analysis, theoretically it is possible to transform to $\phi_1^2, \ldots, \phi_{p-1}^2$ from $w_1, w_2, \ldots, w_{p-1}$ using (5·12) and then to integrate out the other irrelevant co-ordinates of **W** in (6·5) and thus derive the conditional distribution of θ_i^2 and ϕ_j^2 when λ^2 is fixed. However, this does not seem to be

possible, in the general case, as $f_2(\mathbf{W})$ is not known. In the particular case $p = 2$, however, it can be easily shown that

$$f_2(\mathbf{W}) = \text{const.} \frac{1}{\sqrt{(1-w^2)}}, \tag{6·6}$$

where $$\mathbf{W} = \begin{bmatrix} w & \sqrt{(1-w^2)} \\ -\sqrt{(1-w^2)} & w \end{bmatrix}.$$

Hence the joint distribution of θ_1^2 and θ_2^2 when λ^2 is fixed is

$$\frac{\text{const.}\,(\theta_1^2\theta_2^2)^{\frac{1}{2}(n-3)} \exp\{-\frac{1}{2}(\theta_1^2+\theta_2^2)\}\,(\theta_1^2-\theta_2^2)\,d\theta_1^2\,d\theta_2^2}{e^{-\frac{1}{2}\lambda^2}\,(\lambda^2)^{\frac{1}{2}n-1}\,(\theta_1^2-\lambda^2)^{\frac{1}{2}}\,(\lambda^2-\theta_2^2)^{\frac{1}{2}}}. \tag{6·7}$$

When $p = 2$, we find from (5·11)

$$\begin{aligned} \chi_d^2 &= \theta_1^2+\theta_2^2-\lambda^2-\theta_1^2\theta_2^2/\lambda^2 \\ &= (\theta_1^2-\lambda^2)\,(\lambda^2-\theta_2^2)/\lambda^2. \end{aligned} \tag{6·8}$$

In (6·7), make the transformation

$$t_{21}^2 = (\theta_1^2-\lambda^2)\,(\lambda^2-\theta_2^2)/\lambda^2 \tag{6·9}$$

$$t_{22}^2 = \theta_1^2\theta_2^2/\lambda^2. \tag{6·10}$$

The Jacobian of transformation is $\lambda^2/(\theta_1^2-\theta_2^2)$, and hence it will be found that t_{21}^2 and t_{22}^2 are independently distributed as χ^2 with 1 and $(n-1)$ degrees of freedom, respectively. One can readily identify these t_{21} and t_{22} with the elements of the matrix $\mathbf{T}$ in (4·2). As Williams (1952) has observed in canonical analysis, t_{21}^2 is dependent mainly on θ_1^2, while t_{22}^2 depends mainly on θ_2^2, even though both are formally symmetric functions of θ_1^2 and θ_2^2. This is so because if the hypothetical principal component is identical with the true one, λ^2 must lie in the neighbourhood of θ_1^2 and hence

$$\chi_d^2 = t_{21}^2 \sim (\theta_1^2-\lambda^2)/\lambda^2,$$

$$\chi^2-\chi_d^2 = t_{22}^2 \sim \theta_2^2$$

and thus χ_d^2 is appropriate for testing the concordance of the data with the hypothetical principal component, while $\chi^2-\chi_d^2$ gives a test for departure from the assumption of adequacy of a single non-isotropic principal component. Williams (1952) calls this latter test a 'fluid' one because it depends on the choice of the hypothetical principal component and the resulting λ^2. The justification for the use of t_{22}^2 is that it is independent of the unknown parameter σ_1^2, being dependent instead on a statistic which, while to some extent arbitrary, is yet determinate.

7. ILLUSTRATION

To illustrate the goodness-of-fit test, derived in this paper, a random sample of size $n = 50$ was drawn, using Wold's (1948) random normal deviates, from a multivariate normal distribution of four variables x_1, x_2, x_3, x_4 of zero means and variance-covariance matrix

$$\mathbf{\Sigma} = (\sigma_1^2-1)\,\mathbf{l}\mathbf{l}' + \mathbf{I},$$

where $\sigma_1^2 \geqslant 1$, $\mathbf{l}'\mathbf{l} = 1$ and $\mathbf{l}'$ was taken proportional to

$$\mathbf{m}' = [m_1, m_2, m_3, m_4] = [0{\cdot}3333,\ -0{\cdot}6666, 1{\cdot}0, 0{\cdot}3333].$$

The matrix A of the sum of squares and products of the sample values was

$$\begin{bmatrix} 60{\cdot}3282 & -43{\cdot}6958 & 74{\cdot}0990 & 27{\cdot}1455 \\ & 124{\cdot}7021 & -120{\cdot}1149 & -33{\cdot}5915 \\ & & 218{\cdot}3674 & 62{\cdot}0241 \\ & & & 76{\cdot}2765 \end{bmatrix}.$$

The latent roots of $\mathbf{A}/n$, the estimate of $\boldsymbol{\Sigma}$ are

$$\theta_1' = 6{\cdot}9748, \quad \theta_2' = 1{\cdot}1783, \quad \theta_3' = 0{\cdot}8194, \quad \theta_4' = 0{\cdot}6211.$$

The assumption that $\boldsymbol{\Sigma}$ has only one root > 1 and all others all equal to unity can be tested, if desired, by using Bartlett's statistic (1950, see also Lawley, 1956)

$$\left[(n-1)-\frac{1}{6}\left(2p-1-\frac{2}{p}\right)-\frac{1}{p}\right][-\log_e(\theta_2'\theta_3'\theta_4')+\theta_2'+\theta_3'+\theta_4'-(p-1)].$$

In the example under consideration, the value of this is 6·2157. Comparing with the 5% value of a χ^2 with $p(p-1)/2 = 6$ degrees of freedom, this is not significant. There is thus no evidence against the adequacy of a single non-isotropic principal component. We shall now test the null-hypothesis that the single non-isotropic principal component is $\mathbf{l}'$, proportional to

$$\mathbf{m}' = [0{\cdot}3333, \quad -0{\cdot}6666, \quad 1{\cdot}0, \quad 0{\cdot}3333]$$

(since the example is constructed artificially we know that this is the true principal component). The over-all χ^2 of § 2 gives

$$\begin{aligned} \chi^2 &= tr\mathbf{A}-\mathbf{l}'\mathbf{A}\mathbf{l} \\ &= tr\mathbf{A}-\frac{\mathbf{m}'\mathbf{A}\mathbf{m}}{\mathbf{m}'\mathbf{m}} \\ &= 131{\cdot}522 \end{aligned}$$

for $n(p-1) = 150$ degrees of freedom and is not significant. Its direction component is

$$\begin{aligned} \chi_d^2 &= \frac{\mathbf{l}'\mathbf{A}^2\mathbf{l}}{\mathbf{l}'\mathbf{A}\mathbf{l}}-\mathbf{l}'\mathbf{A}\mathbf{l} \\ &= \frac{\mathbf{m}'\mathbf{A}^2\mathbf{m}}{\mathbf{m}'\mathbf{A}\mathbf{m}}-\frac{\mathbf{m}'\mathbf{A}\mathbf{m}}{\mathbf{m}'\mathbf{m}} \\ &= 0{\cdot}523 \end{aligned}$$

for $(p-1) = 3$ degrees of freedom and is not significant. The hypothetical principal component is thus in agreement with the hypothesis, as it should be. However, if we take any other hypothetical component, say one proportional to

$$[1, \quad 1, \quad -2, \quad 3]$$

we find $\chi^2 = 409{\cdot}0945$ and $\chi_d^2 = 69{\cdot}9746$. Both are significant and χ_d^2 indicates that the proposed principal component differs from the true one significantly in direction. If $\mathbf{m}$ is the latent vector of $\boldsymbol{\Sigma}$, corresponding to the largest root, it is obvious that only the ratio $m_1:m_2:m_3:m_4$ is unique, and hence any one of them can be taken to be unity. Let us choose $m_3 = 1$ and then the sample estimates of the three ratios m_1, m_2, m_4 are provided by the first latent vector of $\mathbf{A}$, which in the present example is

$$\hat{\mathbf{m}}' = [0{\cdot}3904, \quad -0{\cdot}6645, \quad 1{\cdot}0, \quad 0{\cdot}3485].$$

Bartlett's (1951*a*) method of substituting the estimates of all of m_1, m_2, m_3, m_4 but one in χ^2 and equating it to the 5 % point of the χ^2 distribution with appropriate degrees of freedom [$n(p-1)-2$ and not $n(p-1)$ as estimates of two ratios are substituted] will yield the approximate 95 % confidence intervals for m_1, m_2, m_3 and m_4 readily [$m_3 = 1$ but to preserve symmetry, a confidence interval for that also was obtained by Bartlett]. These are given at the end of this section. This method can be used with χ_a^2 also, but in this case it is not straightforward, because the inequality $\chi_a^2 \leqslant 5$ % pt. of a χ^2 with $3-2=1$ degree of freedom, where

$$\chi_a^2 = \frac{\mathbf{m'A^2m}}{\mathbf{m'Am}} - \frac{\mathbf{m'Am}}{\mathbf{m'm}},$$

yields successively,

$$g_1(m_1) = m_1^4 + 7{\cdot}3025m_1^3 + 9{\cdot}9037m_1^2 - 11{\cdot}4221m_1 + 2{\cdot}1266 \leqslant 0,$$

$$g_2(m_2) = m_2^4 - 2{\cdot}4932m_2^3 - 1{\cdot}1106m_2^2 + 3{\cdot}3603m_2 + 1{\cdot}5652 \leqslant 0,$$

$$g_3(m_3) = m_3^4 - 0{\cdot}6342m_3^3 - 1{\cdot}4439m_3^2 + 0{\cdot}3897m_3 + 0{\cdot}5163 \leqslant 0,$$

$$g_4(m_4) = m_4^4 + 8{\cdot}6774m_4^3 + 14{\cdot}8688m_4^2 - 14{\cdot}2838m_4 + 2{\cdot}0636 \leqslant 0,$$

when estimates of all of m_1, m_2, m_3, m_4 but one are substituted in χ_a^2. The roots of $g_1(m_1) = 0$ are

$$0{\cdot}2516, \quad 0{\cdot}5264, \quad -4{\cdot}5558, \quad -3{\cdot}5248$$

and as $g_1(m_1) \leqslant 0$, the confidence interval for m_1 is either (0·2516, 0·5264) or (−4·5558, −3·5248); but as the sample estimate of m_1 is 0·3904, it is reasonable to take the confidence interval (0·2516, 0·5264). Similarly, the roots of $g_2(m_2) \equiv 0$ are

$$-0{\cdot}8646, \quad -0{\cdot}4914, \quad 2{\cdot}0663, \quad 1{\cdot}7829$$

and as $g_2(m_2) \leqslant 0$ and $\hat{m}_2 = -0{\cdot}6645$, the confidence interval for m_2 is (−0·8646, −0·4914). The roots of the other two quartic equations $g_3(m_3) = 0$ and $g_4(m_4) = 0$ are respectively

$$0{\cdot}7899, \quad 1{\cdot}2781, \quad -0{\cdot}7654, \quad -0{\cdot}6684$$

and

$$0{\cdot}1832, \quad 0{\cdot}5256, \quad -5{\cdot}4675, \quad -3{\cdot}9186.$$

Arguing as in the case of m_1 and m_2, the confidence intervals for m_3 and m_4 are (0·7899, 1·2781) and (0·1832, 0·5256), respectively. Thus we get the following table:

Parameter	True value	Sample estimate	Confidence interval: Based on χ^2	Confidence interval: Based on χ_a^2
m_1	0·3333	0·3904	(−0·1118, 1·0414)	(0·2516, 0·5264)
m_2	−0·6666	−0·6645	(−1·5262, −0·1479)	(−0·8646, −0·4914)
m_3	1·0	1·0	(0·4342, 2·6623)	(0·7899, 1·2781)
m_4	0·3333	0·3485	(0·1762, 1·0153)	(0·1832, 0·5256)

It can thus be seen that the confidence intervals based on χ_a^2 are much narrower than those based on the overall χ^2 and provide much better information, though the calculations require solving quartic equations.

My sincere thanks are due to Professor M. S. Bartlett for his kind interest in this work and for much valuable guidance and helpful discussions during the preparation of this paper.

References

Anderson, T. W. (1958). *Introduction to Multivariate Statistical Analysis.* New York: John Wiley and Sons Inc.

Bartlett, M. S. (1947). The general canonical correlation distribution. *Ann. Math. Statist.* **18**, 1–17.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *Brit. J. Psych.* (Stat. Sec.) **3**, 77–85.

Bartlett, M. S. (1951*a*). The goodness of fit of a single hypothetical discriminant function in the case of several groups. *Ann. Eugen., Lond.*, **16**, 199–214.

Bartlett, M. S. (1951*b*). The effect of standardization on a χ^2-approximation in factor analysis. *Biometrika*, **38**, 337–44.

Bartlett, M. S. (1951*c*). A further note on tests of significance in factor analysis. *Brit. J. Psych.* (Stat. Sec.) **4**, 1–2.

Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2-approximations. *J. R. Statist. Soc.* B, **16**, 296–8.

Deemer, W. L. & Olkin (1951). The Jacobians of certain matrix transformations useful in multivariate analysis. *Biometrika*, **38**, 345–67.

Halmos, P. R. (1950). *Measure Theory.* New York: D. Van Nostrand.

Kshirsagar, A. M. (1960). A note on the derivation of some exact multivariate tests. *Biometrika*, **47**, 480–2.

Lawley, D. N. (1953). A modified method of estimation in factor analysis and some large sample results. *Uppsala Symp. on Psychological Factor Analysis*, 17–19 March, 1953, pp. 35–42. Uppsala: Almqvist and Wiksell.

Lawley, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128–36.

Mauldon, J. G. (1955). Pivotal quantities for Wishart's and related distributions and a paradox in fiducial theory. *J.R. Statist. Soc.* B, **17**, 79–85.

Williams, E. J. (1952). Some exact tests in multivariate analysis. *Biometrika* **49**, 17–31.

Williams, E. J. (1955). Significance tests for discriminant functions and linear functional relationships. *Biometrika*, **42**, 360–81.

Wold, H. (1948). *Tracts for Computers*, no. 25. Cambridge University Press.

Biometrika (1961), **48**, 3 *and* 4, *p.* 409
Printed in Great Britain

On the evaluation of the probability integral of the multivariate *t*-distribution

BY S. JOHN
Indian Statistical Institute, Calcutta

1. INTRODUCTION AND SUMMARY

A random vector $t = (t_1, t_2, \ldots, t_p)$ is said to have the p-variate t-distribution with n degrees of freedom if its distribution in p-dimensional Euclidean space has for its density function the function

$$g_n(t_1, t_2, \ldots, t_p; \mathrm{P}) = \frac{|A|^{\frac{1}{2}}\,\Gamma(\frac{1}{2}[n+p])}{(n\pi)^{\frac{1}{2}p}\,\Gamma(\frac{1}{2}n)}\left[1+\frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{p}\alpha_{ij}t_it_j\right]^{-\frac{1}{2}(n+p)}$$
$$(-\infty < t_i < \infty;\ i = 1, 2, \ldots, p), \qquad (1)$$

with $\mathrm{P} = (\rho_{ij})$ a positive definite $p \times p$ matrix having every one of its diagonal elements equal to unity, $A = (\alpha_{ij}) = \mathrm{P}^{-1}$, and $|A|$ denoting the determinant of the matrix A. We shall introduce the symbol $G_n(h_1, h_2, \ldots, h_p; \mathrm{P})$ to denote the corresponding distribution function, i.e.

$$G_n(h_1, h_2, \ldots, h_p; \mathrm{P}) = \int_{-\infty}^{h_1}\int_{-\infty}^{h_2}\cdots\int_{-\infty}^{h_p} g_n(t_1, \ldots, t_p; \mathrm{P})\,dt_1 \ldots dt_p. \qquad (2)$$

In the bivariate case, we shall write $g_n(t_1, t_2; \rho_{12})$ for $g_n(t_1, t_2; \mathrm{P})$ and $G_n(t_1, t_2; \rho_{12})$ for $G_n(t_1, t_2; \mathrm{P})$.

If the random variables $x_1, \ldots, x_p$ follow the multivariate normal law with means zero, common variance σ^2 and a correlation matrix (ρ_{ij}) and if $(ns^2)/\sigma^2$ is an independent χ^2 variable with n degrees of freedom, the random vector $t = (t_1, \ldots, t_p)$, where $t_i = x_i/s\,(i = 1, \ldots, p)$ will have $g_n(t_1, \ldots, t_p; \mathrm{P})$ as density function. Bechhofer, Dunnett & Sobel (1954) consider this distribution in connexion with a problem in the ranking of means of normal populations. Dunnett & Sobel (1954) give a formula for evaluating the probability integral when $p = 2$. Using this, they have prepared tables of the function $G_n(h, h;\ \pm 0{\cdot}5)$ and its inverse. Dunnett & Sobel (1955) provide approximations for $G_n(h_1, \ldots, h_p; \mathrm{P})$, valid when P satisfies certain conditions. In this paper we give an alternative formula for the evaluation of the probability integral. Though we too discuss only the bivariate case in detail, our method is of wider applicability in the sense that it can be adopted to get the probability integral of the multivariate t-distribution of any dimension.

We must mention here that the method of this paper is similar to that of Kendall (1941) for evaluating the probability integral of the multivariate normal distribution.

We have already mentioned that the multivariate t-distribution arises in the ranking of normal populations according to their means. We give more applications towards the end of this paper. It is shown how the multivariate t-distribution can be used in setting up simultaneous confidence bounds for the means of correlated normal variables. Other applications are in constructing simultaneous confidence bounds for the parameters in a linear model and for future observations from a multivariate normal distribution. Further applications will appear in a later paper.

2. The characteristic function

To derive our formula for the probability integral we require the characteristic function. For this purpose, we may, without loss of generality, put $\sigma = 1$. The characteristic function is given by

$$\begin{aligned}\phi(\theta_1, \theta_2, \ldots, \theta_p) &= \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \int_0^{\infty} \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}} \exp\left(i\theta_1 x_1/s + i\theta_2 x_2/s + \ldots + i\theta_p x_p/s\right) \\ &\quad \times \exp\left(-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\alpha_{ij}x_i x_j\right) \frac{(ns^2)^{\frac{1}{2}n-1}}{2^{\frac{1}{2}n-1}\Gamma(\frac{1}{2}n)} e^{-\frac{1}{2}ns^2} ns\, ds\, dx_1 \ldots dx_p \\ &= n^{\frac{1}{2}n}\int_0^{\infty} \frac{s^{n+p-1}}{2^{\frac{1}{2}n-1}\Gamma(\frac{1}{2}n)} e^{-\frac{1}{2}ns^2} ds \left[\int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}} \exp\left(i\theta_1 z_1 + \ldots + i\theta_p z_p\right)\right. \\ &\quad \left. \times \exp\left(-\tfrac{1}{2}s^2 \sum_i \sum_j \alpha_{ij} z_i z_j\right) dz_1 \ldots dz_p\right] \\ &= \int_0^{\infty} \frac{(ns^2)^{\frac{1}{2}n-1}}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} e^{-\frac{1}{2}ns^2} d(ns^2) \exp\left(-\frac{1}{2s^2}\theta \mathrm{P} \theta'\right) \\ &= \frac{1}{\Gamma(\frac{1}{2}n)}\int_0^{\infty} t^{\frac{1}{2}n-1} \exp\left(-t - \frac{n}{4t}\theta \mathrm{P}\theta'\right) dt, \end{aligned} \tag{3}$$

where

$$\theta = (\theta_1, \ldots, \theta_p).$$

3. The probability integral

For the sake of simplicity we now restrict our attention to the case $p = 2$; exactly similar methods will give the probability integral whatever positive integral value p has.

In this case,

$$\phi(\theta_1, \theta_2) = \frac{1}{\Gamma(\frac{1}{2}n)}\int_0^{\infty} t^{\frac{1}{2}n-1} \exp\left[-t - \frac{n}{4t}(\theta_1^2 + \theta_2^2)\right]\left[\sum_{r=0}^{\infty}\frac{1}{r!}\left(-\frac{n\rho}{2t}\right)^r \theta_1^r \theta_2^r\right] dt, \tag{4}$$

where we have set $\rho = \rho_{12}$.

By the inversion theorem, the joint density function $g_n(t_1, t_2; \rho)$ of t_1 and t_2 is given by

$$\begin{aligned} g_n(t_1, t_2; \rho) &= \frac{1}{(2\pi)^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\left(-i\theta_1 t_1 - i\theta_2 t_2\right) d\theta_1 d\theta_2 \\ &\quad \times \frac{1}{\Gamma(\frac{1}{2}n)}\int_0^{\infty} t^{\frac{1}{2}n-1}\exp\left[-t - \frac{n}{4t}(\theta_1^2+\theta_2^2)\right] dt \left[\sum_{r=0}^{\infty}\frac{1}{r!}\left(-\frac{n\rho}{2t}\right)^r \theta_1^r \theta_2^r\right] \\ &= \frac{1}{\Gamma(\frac{1}{2}n)}\int_0^{\infty} e^{-t} t^{\frac{1}{2}n-1} dt \left[\sum_{r=0}^{\infty}\frac{1}{r!}\left(\frac{n\rho}{2t}\right)^r \left(\frac{\partial}{\partial t_1}\right)^r \left(\frac{t}{n\pi}\right)^{\frac{1}{2}}\right. \\ &\quad \left. \times \exp\left(-\frac{1}{n}t_1^2 t\right)\left(\frac{\partial}{\partial t_2}\right)^r \left(\frac{t}{n\pi}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{n}t_2^2 t\right)\right]. \end{aligned} \tag{5}$$

We are now in a position to evaluate the probability integral. It is easy to see that

$$G_n(h_1, h_2; \rho) = y_{n,0}(h_1, h_2) + \frac{1}{2\pi}\sum_{r=1}^{\infty}\frac{\rho^r}{r!} y_{n,r}(h_1, h_2), \tag{6}$$

where

$$\begin{aligned} y_{n,r}(h_1, h_2) &= \frac{2}{n\Gamma(\frac{1}{2}n)}\int_0^{\infty} e^{-t} t^{\frac{1}{2}n} dt \left(\frac{n}{2t}\right)^r \left[\left(\frac{\partial}{\partial h_1}\right)^{r-1} \exp\left(-\frac{1}{n}h_1^2 t\right)\right]\left[\left(\frac{\partial}{\partial h_2}\right)^{r-1} \exp\left(-\frac{1}{n}h_2^2 t\right)\right] \\ &= \frac{1}{\Gamma(\frac{1}{2}n)}\int_0^{\infty} \exp\left[-t\left\{1 + \frac{1}{n}(h_1^2 + h_2^2)\right\}\right] t^{\frac{1}{2}n-1} dt H_{r-1}\left(\left[\frac{2t}{n}\right]^{\frac{1}{2}} h_1\right) H_{r-1}\left(\left[\frac{2t}{n}\right]^{\frac{1}{2}} h_2\right) \\ &\qquad (r = 1, 2, \ldots). \end{aligned} \tag{7}$$

Here $H_r(x)$ denotes the Hermite polynomial of degree r defined by

$$\left(-\frac{d}{dx}\right)^r e^{-\frac{1}{2}x^2} = H_r(x)\, e^{-\frac{1}{2}x^2}, \tag{8}$$

and

$$\begin{aligned} y_{n,0}(h_1, h_2) &= \frac{(n\pi)^{-1}}{\Gamma(\frac{1}{2}n)} \int_0^\infty e^{-t} t^{\frac{1}{2}n} \left[\int_{-\infty}^{h_1} \exp\left(-\frac{1}{n} t_1^2 t\right) dt_1\right]\left[\int_{-\infty}^{h_2} \exp\left(-\frac{1}{n} t_2^2 t\right) dt_2\right] dt \\ &= \frac{1}{\Gamma(\frac{1}{2}n)} \int_0^\infty e^{-t} t^{\frac{1}{2}n-1}\, dt\, G\left(\left[\frac{2t}{n}\right]^{\frac{1}{2}} h_1\right) G\left(\left[\frac{2t}{n}\right]^{\frac{1}{2}} h_2\right), \end{aligned} \tag{9}$$

where

$$G(x) = \int_{-\infty}^{x} \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}u^2}\, du. \tag{10}$$

We give below explicit expressions for $y_{n,r}(h_1, h_2)$ $(r = 1, 2, \ldots, 6)$.

If we set

$$z = 1 + (1/n)\,(h_1^2 + h_2^2), \tag{11}$$

$$y_{n,1}(h_1, h_2) = z^{-\frac{1}{2}n}, \tag{12}$$

$$y_{n,2}(h_1, h_2) = h_1 h_2 z^{-(\frac{1}{2}n+1)}, \tag{13}$$

$$y_{n,3}(h_1, h_2) = \{1 + (2/n)\}\, h_1^2 h_2^2 z^{-(\frac{1}{2}n+2)} - (h_1^2 + h_2^2)\, z^{-(\frac{1}{2}n+1)} + z^{-\frac{1}{2}n}, \tag{14}$$

$$y_{n,4}(h_1, h_2) = \left(1+\frac{2}{n}\right)\left(1+\frac{4}{n}\right) h_1^3 h_2^3 z^{-(\frac{1}{2}n+3)} - 3\left(1+\frac{2}{n}\right) h_1 h_2 (h_1^2 + h_2^2)\, z^{-(\frac{1}{2}n+2)} + 9 h_1 h_2 z^{-(\frac{1}{2}n+1)}, \tag{15}$$

$$\begin{aligned} y_{n,5}(h_1, h_2) = {} & \left(1+\frac{2}{n}\right)\left(1+\frac{4}{n}\right)\left(1+\frac{6}{n}\right) h_1^4 h_2^4 z^{-(\frac{1}{2}n+4)} - 6\left(1+\frac{2}{n}\right)\left(1+\frac{4}{n}\right) h_1^2 h_2^2 (h_1^2 + h_2^2)\, z^{-(\frac{1}{2}n+3)} \\ & + 3\left(1+\frac{2}{n}\right)(h_1^4 + 12 h_1^2 h_2^2 + h_2^4)\, z^{-(\frac{1}{2}n+2)} - 18(h_1^2 + h_2^2)\, z^{-(\frac{1}{2}n+1)} + 9 z^{-\frac{1}{2}n}, \end{aligned} \tag{16}$$

$$\begin{aligned} y_{n,6}(h_1, h_2) = {} & \left(1+\frac{2}{n}\right)\left(1+\frac{4}{n}\right)\left(1+\frac{6}{n}\right)\left(1+\frac{8}{n}\right) h_1^5 h_2^5 z^{-(\frac{1}{2}n+5)} \\ & - 10\left(1+\frac{2}{n}\right)\left(1+\frac{4}{n}\right)\left(1+\frac{6}{n}\right) h_1^3 h_2^3 (h_1^2 + h_2^2)\, z^{-(\frac{1}{2}n+4)} \\ & + 5\left(1+\frac{2}{n}\right)\left(1+\frac{4}{n}\right) h_1 h_2 (3 h_1^4 + 20 h_1^2 h_2^2 + 3 h_2^4)\, z^{-(\frac{1}{2}n+3)} \\ & - 150\left(1+\frac{2}{n}\right) h_1 h_2 (h_1^2 + h_2^2)\, z^{-(\frac{1}{2}n+2)} + 225 h_1 h_2 z^{-(\frac{1}{2}n+1)}. \end{aligned} \tag{17}$$

The evaluation of $y_{n,0}(h_1, h_2)$ requires separate consideration.

4. Evaluation of $y_{n,0}(h_1, h_2)$

First we observe that

$$y_{n,0}(h_1, h_2) = G_n(h_1, h_2;\ 0). \tag{18}$$

Equation (9) can be used to evaluate $y_{n,0}(h_1, h_2)$. The integration involved has to be done numerically. Gauss's formula for numerical quadrature (Kopal, 1955, p. 371) will be found especially convenient. By this method we have prepared tables of $y_{n,0}(h_1, h_2)$ for $n = 11, 12$. The first of these, that for $n = 11$, is reproduced at the end of this paper as Table 2. The next section gives recurrence relations connecting these values of n with other values of n.

Other methods for the evaluation of $y_n(h_1, h_2)$ are given by John (1961). The table was computed on HEC–2M at the Indian Statistical Institute. The programme was prepared by

Mr P. K. Mitra, and Mr B. Mukherjee was responsible for the running of it off the computer. A similar table for the case $n = 12$ is available at the Institute.

In the tables, $y_{n,0}(h_1, h_2)$ has been tabulated only for values of h_1 and h_2 satisfying the inequalities $h_1 \geqslant 0$, $h_2 \geqslant 0$, $h_2 \geqslant h_1$. Values of $y_{n,0}(h_1, h_2)$ for h_1 or h_2 negative can be found from the formulae

$$y_{n,0}(h_1, h_2) = G_n(h_2) - y_{n,0}(-h_1, h_2), \tag{19}$$

$$y_{n,0}(h_1, h_2) = G_n(h_1) - y_{n,0}(h_1, -h_2), \tag{20}$$

where $G_n(x)$ is the probability that a random variable having Student's t-distribution with n degrees of freedom has a value less than or equal to x. If both h_1 and h_2 are negative, we may use the formula

$$y_{n,0}(h_1, h_2) = 1 + y_{n,0}(-h_1, -h_2) - G_n(-h_1) - G_n(-h_2). \tag{21}$$

Further, there is no loss of generality in assuming $h_2 \geqslant h_1$ since

$$y_{n,0}(h_1, h_2) = y_{n,0}(h_2, h_1). \tag{22}$$

5. Some useful recurrence relations

From equation (9), we get by integration by parts the following recurrence relation:

$$y_{n,0}(h_1, h_2) = y_{n+2,0}\left(\left[1+\frac{2}{n}\right]^{\frac{1}{2}} h_1, \left[1+\frac{2}{n}\right]^{\frac{1}{2}} h_2\right) - (n^3\pi)^{-\frac{1}{2}} \frac{\Gamma(\frac{1}{2}n+\frac{1}{2})}{\Gamma(\frac{1}{2}n)} \left[\frac{h_1 G_{n+1}\left(\left[\frac{n+1}{n+h_1^2}\right]^{\frac{1}{2}} h_2\right)}{\{1+(1/n)h_1^2\}^{\frac{1}{2}(n+1)}} + \frac{h_2 G_{n+1}\left(\left[\frac{n+1}{n+h_2^2}\right]^{\frac{1}{2}} h_1\right)}{\{1+(1/n)h_2^2\}^{\frac{1}{2}(n+1)}}\right]. \tag{23}$$

In getting equation (23), $t^{\frac{1}{2}n-1}$ was the factor which we selected for the first integration. If we now select the factor e^{-t} for the first integration, we get the recurrence relation given below:*

$$y_{n,0}(h_1, h_2) = y_{n-2,0}\left(\left[1-\frac{2}{n}\right]^{\frac{1}{2}} h_1, \left[1-\frac{2}{n}\right]^{\frac{1}{2}} h_2\right) + \tfrac{1}{2}(n\pi)^{-\frac{1}{2}} \frac{\Gamma(\frac{1}{2}n-\frac{1}{2})}{\Gamma(\frac{1}{2}n)} \left[\frac{h_1 G_{n-1}\left(\left[\frac{n-1}{n+h_1^2}\right]^{\frac{1}{2}} h_2\right)}{\{1+(1/n)h_1^2\}^{\frac{1}{2}(n-1)}} + \frac{h_2 G_{n-1}\left(\left[\frac{n-1}{n+h_2^2}\right]^{\frac{1}{2}} h_1\right)}{\{1+(1/n)h_2^2\}^{\frac{1}{2}(n-1)}}\right] \quad (n > 2). \tag{24}$$

By the same procedure we can express $y_{2,0}(h_1, h_2)$ entirely in terms of the probability integral of the univariate t-distribution. This relationship is given by the following equation:

$$y_{2,0}(h_1, h_2) = \tfrac{1}{4} + 8^{-\frac{1}{2}} \left[\frac{h_1 G_1([2+h_1^2]^{-\frac{1}{2}} h_2)}{(1+\frac{1}{2}h_1^2)^{\frac{1}{2}}} + \frac{h_2 G_1([2+h_2^2]^{-\frac{1}{2}} h_1)}{(1+\frac{1}{2}h_2^2)^{\frac{1}{2}}}\right]. \tag{25}$$

This, together with (24), shows that $y_{n,0}(h_1, h_2)$ for all even n can be built up from tables of the probability integral of the univariate t-distribution. If n is a small even integer this procedure is to be preferred. Similarly if n is a fairly large integer (odd or even) it may be advantageous to connect it with the probability integral of the bivariate normal distribution through the formula (23).

* Relation (24) can be derived also from (23).

5. Comparisons with the values given by Bechhofer, Dunnett and Sobel (1954)

As a check on the correctness of our formula and as a means of seeing how good an approximation is obtained by taking the sum of a few terms from the beginning in the series expansion (6), we carried out some computations. The results are presented in Table 1, where we have taken $n = 12$ and $h_1 = h_2 = h$; in the approximation, we have used the first six terms in summing the second expression on the right-hand side of equation (6). It is seen that for $\rho = \pm 0{\cdot}5$, the sum of the first seven terms of the series expansion (6) provides an approximation to the correct value which is satisfactory for most purposes. If the value of $|\rho|$ is larger, more terms have to be included in the summation to get the same accuracy. We have developed simpler methods for computing the probability integral in such cases. These methods will be published when the required tables are ready.

Table 1. *Values of* $G_{12}(h, h; \pm 0{\cdot}5)$

	Case $\rho = 0{\cdot}5$, $n = 12$		Case $\rho = -0{\cdot}5$, $n = 12$	
h	Approx. from (6)	Exact value	Approx. from (6)	Exact value
0·00	0·3333	0·33333	0·1667	0 16667
·25	·4355	·43555	·2799	·27988
·50	·5421	·54150	·4131	·41366
·75	·6429	·64292	·5488	·54880
1·00	·7330	·73301	·6694	·66936

6. The multivariate case

We have discussed in detail how the probability integral of the bivariate t-distribution may be evaluated. It is possible to get an expression similar to (6) for the probability integral of a multivariate t-distribution of any dimension by methods similar to those which were adopted in the bivariate case. The formula is the following:

$$G_n(h_1, h_2, \ldots, h_p; \mathrm{P}) = G_n(h_1, h_2, \ldots, h_p; I) + (2\pi)^{-\frac{1}{2}p} \sum_{r=1}^{\infty} \frac{1}{r!} Y_{n,r}(h_1, \ldots, h_p; \mathrm{P}), \tag{26}$$

where

$$Y_{n,r}(h_1, \ldots, h_p; \mathrm{P}) = \frac{1}{\Gamma(\frac{1}{2}n)} \int_0^{\infty} t^{\frac{1}{2}n-1} \exp\left[-t\{1 + (1/n)(h_1^2 + \ldots + h_p^2)\}\right] U_{n,r}(t; h_1, \ldots, h_p; \mathrm{P})\, dt, \tag{27}$$

and

$$U_{n,r}(t; h_1, \ldots, h_p; \mathrm{P}) = \left\{\prod_{i=1}^{p} H^{-1}\left(\left[\frac{2t}{n}\right]^{\frac{1}{2}} h_i\right)\right\}\left[\sum_{i<j}^{p} \rho_{ij} H\left(\left[\frac{2t}{n}\right]^{\frac{1}{2}} h_i\right) H\left(\left[\frac{2t}{n}\right]^{\frac{1}{2}} h_j\right)\right]^r. \tag{28}$$

It is to be understood that in (28) after expansion of the integrand, $H^r([2t/n]^{\frac{1}{2}} h_i)$ is to be replaced by $H_r([2t/n]^{\frac{1}{2}} h_i)$ if $r \geqslant 0$ and by $(2\pi)^{\frac{1}{2}} G([2t/n]^{\frac{1}{2}} h_i) \exp\{(1/n) t h_i^2\}$ if $r = -1$.

For many of the terms in (26), explicit expressions similar to those in equations (12) to (17) can be given; but others have to be evaluated by numerical quadrature.

7. Applications

The multivariate t-distribution arises in many statistical problems. In this section we shall describe a few of these problems.

7·1. *Simultaneous confidence bounds for the means of correlated normal variables*

Suppose $x = (x_1, \ldots, x_p)$ follows a multivariate normal distribution with mean vector $\mu = (\mu_1, \ldots, \mu_p)$ and dispersion matrix $\sigma^2 \mathrm{P}$. P is the correlation matrix and we shall suppose that this is known. The parameters μ and σ^2 are unknown. It is desired to set up simultaneous confidence bounds for the μ_i's.

This can be done as follows: draw a sample of size N from the population of x's. Let the observations be x_{ij} $(i = 1, 2, \ldots, p; j = 1, 2, \ldots, N)$, the first subscript pertaining to the variate and the second subscript pertaining to the sample. Let

$$\bar{x}_i = \left(\sum_{r=1}^{N} x_{ir} \right) \Big/ N \quad (i = 1, 2, \ldots, p). \tag{29}$$

Then

$$s^2 = \left[\sum_{i=1}^{p} \sum_{j=1}^{p} \alpha_{ij} \sum_{r=1}^{N} (x_{ir} - \bar{x}_i)(x_{jr} - \bar{x}_j) \right] \Big/ \{(N-1)p\} \tag{30}$$

is an unbiased estimator of σ^2. The random variable $(N-1)ps^2/\sigma^2$ is distributed independently of the $\bar{x}_i$'s as χ^2 with $(N-1)p$ degrees of freedom. We shall set $n = (N-1)p$. Determine h so that

$$\int_{-h}^{h} \ldots \int_{-h}^{h} g_n(t_1, \ldots, t_p; \mathrm{P})\, dt_1, \ldots, dt_p = \alpha, \tag{31}$$

where α is a positive number between zero and one. Then the inequalities

$$\bar{x}_i - N^{-\frac{1}{2}} hs < \mu_i < \bar{x}_i + N^{-\frac{1}{2}} hs \quad (i = 1, 2, \ldots, p) \tag{32}$$

will be simultaneously satisfied with probability α.

The condition that all the variates have the same variance can be slightly relaxed. It is enough that the ratios of the variances are known. Thus if $V(x_i) = c_i \sigma^2 (i = 1, 2, \ldots, p)$ and the c_i's are known, we have only to consider $y_i = c_i^{-\frac{1}{2}} x_i$ in the place of x_i. This will yield simultaneous confidence bounds for $c_i^{-\frac{1}{2}} \mu_i$'s, which can be readily converted into simultaneous confidence bounds for the μ_i's.

We must refer here to a paper by Olive Jean Dunn (1958) discussing methods of setting simultaneous confidence bounds for the expected values of correlated normal variables. She does not require that the correlation matrix be known. She achieves her result by considering p linear combinations of the observations which have the same expectations as the original variables. The first linear combination is a linear combination of the observations on the variable x_1; the second linear combination is a linear combination of the observations on x_2 and so on. In so far as the linear combinations are not unique, this procedure is unsatisfactory.

7·2. *Simultaneous confidence bounds for future observation*

Here again we will suppose that the population is multivariate normal, that all the variates have equal variances and that the correlation matrix is known. A sample of size N is available and it is required to set up simultaneous confidence bounds for the components of a future observation vector $x = (x_1, \ldots, x_p)$.

Consider the inequalities

$$\bar{x}_i - (1 + N^{-1})^{\frac{1}{2}} hs \leqslant x_i \leqslant \bar{x}_i + (1 + N^{-1})^{\frac{1}{2}} hs \quad (i = 1, 2, \ldots, p), \tag{33}$$

$\bar{x}_i$, h and s having the same meanings as in § 7·1. The probability that these inequalities will hold simultaneously is α. The inequalities (33) will thus serve as simultaneous confidence bounds for the components of x.

As in § 7·1 the condition that all the variates should have the same variance can be replaced by the weaker condition that ratios of the variances should be known.

7·3. *Simultaneous confidence bounds for regression coefficients**

In regression analysis we have N independent normal variables $y_1, y_2, \ldots, y_N$ with common variance σ^2 and expectations given by

$$E(y_r) = \beta_0 + \sum_{i=1}^{p} \beta_i x_{ir} \quad (r = 1, 2, \ldots, N). \tag{34}$$

The x_{ij}'s are known constants. The β's are unknown parameters.

Put

$$S_{ij} = \sum_{r=1}^{N} (x_{ir} - \bar{x}_i)(x_{jr} - \bar{x}_j) \quad (i, j = 1, 2, \ldots, p), \tag{35}$$

$$S = (S_{ij}), \tag{36}$$

$$T_{yi} = \sum_{r=1}^{N} y_r (x_{ir} - \bar{x}_i) \quad (i = 1, 2, \ldots, p), \tag{37}$$

$$T_{yy} = \sum_{r=1}^{N} (y_r - \bar{y})^2. \tag{38}$$

Let $(b_1, b_2, \ldots, b_p)$ be a solution of the equations

$$\left.\begin{aligned} S_{11} b_1 + S_{12} b_2 + \ldots + S_{1p} b_p &= T_{y1}, \\ S_{21} b_1 + S_{22} b_2 + \ldots + S_{2p} b_p &= T_{y2}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ S_{p1} b_1 + S_{p2} b_2 + \ldots + S_{pp} b_p &= T_{yp}. \end{aligned}\right\} \tag{39}$$

Also let

$$s^2 = \frac{1}{N-p-1} (T_{yy} - b_1 T_{y1} - b_2 T_{y2} - \ldots - b_p T_{yp}). \tag{40}$$

Then $(b_1, b_2, \ldots, b_p)$ have a multivariate normal distribution with means $(\beta_1, \beta_2, \ldots, \beta_p)$ and variance-covariance matrix $(c_{ij})\,\sigma^2 = S^{-1}\sigma^2$.

s^2 is an unbiased estimator of σ^2. $(N-p-1)\,s^2/\sigma^2$ has a χ^2 distribution with $(N-p-1)$ degrees of freedom and is independent of $(b_1, \ldots, b_p)$. We shall set $n = N-p-1$.

The correlation matrix (ρ_{ij}) for $(b_1, b_2, \ldots, b_p)$ is given by

$$\rho_{ij} = (c_{ii} c_{jj})^{-\frac{1}{2}} c_{ij} \tag{41}$$

and is known. Determine h so that

$$\int_{-h}^{h} \int_{-h}^{h} g_n(t_1, \ldots, t_p; \mathrm{P})\, dt_1 \ldots dt_p = \alpha. \tag{42}$$

Then the probability that the inequalities

$$b_i - h c_{ii}^{\frac{1}{2}} s \leqslant \beta_i \leqslant b_i + h c_{ii}^{\frac{1}{2}} s \quad (i = 1, 2, \ldots, p) \tag{43}$$

are simultaneously satisfied is α. Thus inequalities (43) serve as simultaneous confidence bounds for $\beta_1, \beta_2, \ldots, \beta_p$. Simultaneous confidence bounds for any set of independent linear functions of the parameters can be set up in this fashion. These bounds appear to be more

* I am indebted to Dr C. R. Rao for pointing out the possibility of this application.

relevant than Scheffé's (1953) who gives methods for setting up simultaneous confidence bounds for all possible linear functions of the parameters, standardized in a particular way. Most often we are interested in a few parametric functions and not the whole family of such functions. These functions get a bad deal in Scheffé's method; the width of the confidence interval is unnecessarily large.*

REFERENCES

BECHHOFER, R. E., DUNNETT, C. W. & SOBEL, M. (1954). A two sample multiple decision procedure for ranking means of normal populations with a common variance. *Biometrika*, **41**, 170–6.

DUNN, O. J. (1958). Estimation of the means of dependent variables. *Ann. Math. Statist.* **29**, 1095–111.

DUNNETT, C. W. & SOBEL, M. (1954). A bivariate generalization of Student's t-distribution, with tables for certain special cases. *Biometrika*, **41**, 150–69.

DUNNETT, C. W. & SOBEL, M. (1955). Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution. *Biometrika*, **42**, 258–60.

JOHN, S. (1961). Methods for the evaluation of probabilities of polygonal and angular regions when the distribution is bivariate t. (To be published, probably in *Sankhyā*.)

KENDALL, M. G. (1941). Proof of relations connected with the tetrachoric series and its generalization. *Biometrika*, **32**, 196–8.

KOPAL, Z. (1955). *Numerical Analysis*. London: Chapman and Hall Ltd.

SCHEFFÉ, H. (1953). A method of judging all contrasts in the analysis of variance. *Biometrika*, **40**, 87–104.

Table 2. *Values of* $y_{n,0}(h_1, h_2)$ *for* $n = 11$

h_2 \ h_1	0·0	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
0·0	0·25000									
·1	·26946	0·29051								
·2	·28871	·31133	0·33369							
·3	·30755	·33170	·35558	0·37894						
·4	·32579	·35141	·37675	·40154	0·42552					
0·5	0·34326	0·37029	0·39703	0·42318	0·44848	0·47269				
·6	·35983	·38819	·41624	·44368	·47022	·49562	0·51967			
·7	·37538	·40498	·43427	·46290	·49061	·51712	·54222	0·56575		
·8	·38983	·42059	·45100	·48075	·50953	·53707	·56314	·58759	0·61027	
·9	·40314	·43495	·46641	·49717	·52693	·55541	·58238	·60766	·63111	0·65267
1·0	0·41529	0·44805	0·48045	0·51213	0·54279	0·57212	0·59990	0·62593	0·65010	0·67230
1·2	·43616	·47053	·50452	·53777	·56994	·60072	·62988	·65721	·68258	·70590
1·4	·45272	·48835	·52359	·55806	·59141	·62333	·65357	·68191	·70823	·73243
1·6	·46552	·50210	·53828	·57367	·60791	·64070	·67175	·70088	·72792	·75280
1·8	·47516	·51244	·54932	·58539	·62029	·65372	·68538	·71509	·74267	·76805
2·0	0·48229	0·52008	0·55745	0·59401	0·62940	0·66328	0·69539	0·72551	0·75349	0·77924
2·2	·48747	·52561	·56334	·60025	·63597	·67019	·70261	·73303	·76129	·78731
2·4	·49119	·52958	·56755	·60470	·64067	·67511	·70775	·73838	·76685	·79305
2·6	·49382	·53239	·57053	·60785	·64398	·67858	·71138	·74215	·77070	·79709
2·8	·49568	·53430	·57262	·61006	·64630	·68101	·71391	·74479	·77349	·79991
3·0	0·49698	0·53574	0·57408	0·61160	0·64791	0·68270	0·71567	0·74662	0·77539	0·80187
4·0	·49948	·53838	·57686	·61452	·65097	·68589	·71900	·75008	·77897	·80557
5·0	·49989	·53881	·57732	·61499	·65147	·68641	·71953	·75063	·77953	·80615
6·0	·49997	·53890	·57740	·61508	·65156	·68650	·71963	·75073	·77964	·80626
7·0	·49999	·53892	·57742	·61510	·65158	·68653	·71966	·75076	·77967	·80629
8·0	0·49999	0·53892	0·57743	0·61511	0·65159	0·68653	0·71966	0·75076	0·77967	0·80629

* Scheffé's method is easier to apply, as a referee has pointed out.

Table 2 (*cont.*)

h_2 \ h_1	0·9	1·0	1·2	1·4	1·6	1·8	2·0	2·2	2·4
0·9	0·65267								
1·0	0·67230	0·69253							
1·2	·70590	·72714	0·76351						
1·4	·73243	·75448	·79224	0·82209					
1·6	·75280	·77547	·81430	·84502	0·86864				
1·8	·76805	·79118	·83083	·86222	·88637	0·90450			
2·0	0·77924	0·80271	0·84296	0·87485	0·89939	0·91784	0·93142		
2·2	·78731	·81103	·85172	·88396	·90881	·92749	·94125	0·95122	
2·4	·79305	·81695	·85795	·89045	·91551	·93437	·94827	·95834	0·96554
2·6	·79709	·82111	·86233	·89503	·92024	·93922	·95322	·96337	·97064
2·8	·79991	·82402	·86540	·89822	·92355	·94262	·95669	·96690	·97421
3·0	0·80187	0·82604	0·86753	0·90045	0·92585	0·94499	0·95911	0·96937	0·97671
4·0	·80557	·82985	·87154	·90463	·93019	·94946	·96369	·97403	·98144
5·0	·80615	·83045	·87217	·90530	·93088	·95017	·96442	·97477	·98220
6·0	·80626	·83056	·87229	·90542	·93101	·95030	·96455	·97491	·98234
7·0	·80629	·83059	·87231	·90545	·93104	·95033	·96459	·97495	·98238
8·0	0·80629	0·83059	0·87232	0·90545	0·93104	0·95034	0·96459	0·97495	0·98238

h_2 \ h_1	2·4	2·6	2·8	3·0	4·0	5·0	6·0	7·0	8·0
2·4	0·96554								
2·6	·97064	0·97578							
2·8	·97421	·97939	0·98302						
3·0	0·97671	0·98191	0·98556	0·98812					
4·0	·98144	·98670	·99040	·99299	0·99795				
5·0	·98220	·98747	·99118	·99378	·99876	0·99958			
6·0	·98234	·98762	·99133	·99393	·99892	·99974	0·99990		
7·0	·98238	·98765	·99136	·99396	·99895	·99978	·99994	0·99998	
8·0	0·98238	0·98766	0·99137	0·99397	0·99896	0·99979	0·99995	0·99999	1·00000

Biometrika (1961), **48**, 3 *and* 4, *p.* 419
Printed in Great Britain

Computing the distribution of quadratic forms in normal variables

BY J. P. IMHOF
University of Geneva

1. INTRODUCTION

In this paper exact and approximate methods are given for computing the distribution of quadratic forms in normal variables. In statistical applications the interest centres in general, for a quadratic form Q and a given value x, around the probability $P\{Q > x\}$. Methods of computation have previously been given e.g. by Box (1954), Gurland (1955) and by Grad & Solomon (1955). None of these methods is very easily applicable except, when it can be used, the finite series of Box. Furthermore, all the methods are valid only for quadratic forms in central variables. Situations occur where quadratic forms in non-central variables must be considered as well. Let $\mathbf{x} = (x_1, \ldots, x_n)'$ be a column random vector which follows a multidimensional normal law with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)'$ be a constant vector, and consider the quadratic form $Q = (\mathbf{x}+\boldsymbol{\mu})'\mathbf{A}(\mathbf{x}+\boldsymbol{\mu})$. If $\mathbf{\Sigma}$ is non-singular, one can by means of a non-singular linear transformation (Scheffé (1959), p. 418) express Q in the form

$$Q = \sum_{r=1}^{m} \lambda_r \chi^2_{h_r;\delta_r^2}. \tag{1·1}$$

The λ_r are the distinct non-zero characteristic roots of $\mathbf{A\Sigma}$, the h_r their respective orders of multiplicity, the δ_r are certain linear combinations of $\mu_1, \ldots, \mu_n$ and the $\chi^2_{h_r;\delta_r^2}$ are independent χ^2-variables with h_r d.f. (degrees of freedom) and non-centrality parameter δ_r^2. The variable $\chi^2_{h;\delta^2}$ is defined here by the relation $\chi^2_{h;\delta^2} = (x_1+\delta)^2 + \sum_2^h x_i^2$, where $x_1, x_2, \ldots, x_h$ are independent unit normal deviates. In case $\mathbf{\Sigma}$ is singular but $\mathbf{A}$ is not, a similar decomposition can be obtained except that an additive constant might figure in the right-hand member of (1·1). This is the case for instance if $Q = (x_1+x_2)^2 + x_3^2$, where the linear constraint $x_1+x_2 = c$ is imposed.

We illustrate the reduction of a quadratic form to the form (1·1) on the correlator

$$w = \sum_1^n (x_i - \mu_i)(y_i - \eta_i),$$

where the x_i and y_i are normal deviates with zero means and we assume that

$$\operatorname{cov}(x_i, x_j) = \operatorname{cov}(y_i, y_j) = \delta_{ij}, \quad \operatorname{cov}(x_i, y_j) = \rho\delta_{ij} \quad \text{and} \quad |\rho| < 1.$$

Expressions for the probability density of w have been given in some particular cases of these assumptions by Roe & White (1961). Let

$$\mathbf{z} = (x_1, \ldots, x_n, y_1, \ldots, y_n)' \quad \text{and} \quad \boldsymbol{\zeta} = (\mu_1, \ldots, \mu_n, \eta_1, \ldots, \eta_n)'.$$

Then $w = (\mathbf{z}-\boldsymbol{\zeta})'\mathbf{A}(\mathbf{z}-\boldsymbol{\zeta})$, where $\mathbf{A}$ and the covariance matrix $\mathbf{\Sigma}$ of $\mathbf{z}$ can be partitioned into square submatrices of n rows each: If $\mathbf{U}$ denotes the unit (identity) matrix,

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{U} & \rho\mathbf{U} \\ \rho\mathbf{U} & \mathbf{U} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{U} \\ \mathbf{U} & \mathbf{0} \end{bmatrix}.$$

Under the linear transformation $\mathbf{z}^* = \mathbf{P}\mathbf{z}$, where

$$\mathbf{P} = 2^{-\frac{1}{2}}\begin{bmatrix}(1-\rho)^{-\frac{1}{2}}\mathbf{U} & -(1-\rho)^{-\frac{1}{2}}\mathbf{U}\\ (1+\rho)^{-\frac{1}{2}}\mathbf{U} & (1+\rho)^{-\frac{1}{2}}\mathbf{U}\end{bmatrix},$$

the correlator takes the form $w = (\mathbf{z}^*-\boldsymbol{\zeta}^*)'\mathbf{A}^*(\mathbf{z}^*-\boldsymbol{\zeta}^*)$, where $\boldsymbol{\zeta}^* = \mathbf{P}\boldsymbol{\zeta}$ and $\mathbf{A}^*$ is a diagonal matrix having its n first diagonal elements equal to $\rho-1$, the n remaining ones equal to $\rho+1$. The covariance matrix of $\mathbf{z}^*$ is the unit matrix. The random variable w can therefore be written

$$w = (\rho+1)\chi^2_{n;\delta_1^2} + (\rho-1)\chi^2_{n;\delta_2^2}, \tag{1·2}$$

a linear combination of two independent χ^2-variables with non-centrality parameters $\delta_1^2 = \frac{1}{2}(1+\rho)^{-1}\sum_1^n(\mu_i+\eta_i)^2$ and $\delta_2^2 = \frac{1}{2}(1-\rho)^{-1}\sum_1^n(\mu_i-\eta_i)^2$.

2. FINITE EXPRESSIONS

Box (1954) has considered the case where in (1·1) $\delta_r^2 = 0$ and $h_r = 2\nu_r$ $(r = 1, \ldots, m)$, the ν_r being positive integers. In statistical applications, it is often possible to arrange for the d.f.'s to be even. A problem in the theory of Gaussian noise where this condition is always fulfilled is described by Davenport & Root (1958). Box has then expressed the probability $P\{Q > x\}$ by means of a finite series of χ^2 probabilities. Using a different proof, a modified and often more convenient form of his result will be obtained, as well as a corresponding expression for the probability density $g(x)$ of Q. Suppose $\lambda_1, \ldots, \lambda_m$ have been ordered, so that $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0 > \lambda_{p+1} > \ldots > \lambda_m$. Let $n = \sum_1^m \nu_r$ and $q = \sum_1^p \nu_r$.

THEOREM. *If $x > 0$, then*

$$P\left[\sum_1^m \lambda_r \chi^2_{2\nu_r} > x\right] = \sum_{k=1}^{p} \frac{1}{(\nu_k-1)!}\left[\frac{\partial^{\nu_k-1}}{\partial\lambda^{\nu_k-1}}F_k(\lambda, x)\right]_{\lambda=\lambda_k}, \tag{2·1}$$

where

$$F_k(\lambda, x) = \lambda^{n-1}\exp\{-x/(2\lambda)\}\prod_{\substack{r=1\\ r\neq k}}^{m}(\lambda-\lambda_r)^{-\nu_r}.$$

Proof. Let $\lambda_1', \ldots, \lambda_q', \lambda_{q+1}', \ldots, \lambda_n'$ be real, pairwise unequal parameters such that $\lambda_s' > 0$ for $s = 1, \ldots, q$ and $\lambda_s' < 0$ for $s = q+1, \ldots, n$. Let $Q' = \sum_1^n \lambda_s' \chi_2^2(s)$, where the n χ^2-variables $\chi_2^2(s)$ with two d.f. each are independent. The characteristic function $\phi'(t)$ of Q' is

$$\phi'(t) = \prod_1^n (1-2i\lambda_s' t)^{-1},$$

and correspondingly the probability density is

$$g'(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\prod_1^n(1-2i\lambda_s' t)^{-1}\,dt.$$

The integrand having simple poles only, one finds for $x > 0$

$$g'(x) = \sum_{s=1}^{q} t_s(x)\prod_{s'=1}^{n}{}'(\lambda_s'-\lambda_{s'}')^{-1},$$

where $\prod'$ indicates that the factor corresponding to $s' = s$ is omitted, and where

$$t_s(x) = \tfrac{1}{2}\lambda_s'^{n-2}\exp\{-x/(2\lambda_s')\}.$$

in which X and Y are independent Poisson variables with expected values $\frac{1}{2}x$ and $\frac{1}{2}\delta^2$, respectively. Johnson (1959) rediscovered this equality and used it as a starting point to

The density $g(x)$ of Q can now be obtained by a passage to the limit. Let $\sigma_0 = 0$, $\sigma_k = \sum_{r=1}^{k} \nu_r$ for $k = 1, \ldots, m$. For positive distinct integers $a, b, \ldots, d$ all not larger than m, denote by $\lim_{a,b,\ldots,d}$ the passage to the limit

$$\lambda'_{\sigma_{a-1}+1}, \ldots, \lambda'_{\sigma_a} \to \lambda_a, \ldots, \lambda'_{\sigma_{d-1}+1}, \ldots, \lambda'_{\sigma_d} \to \lambda_d.$$

Then $g(x) = \lim_{1,2,\ldots,m} g'(x)$ can be written

$$\begin{aligned} g(x) = \sum_{k=1}^{p} \lim_{k} \sum_{s=\sigma_{k-1}+1}^{\sigma_k} t_s(x) \prod_{s'=\sigma_{k-1}+1}^{\sigma_k} (\lambda'_s - \lambda'_{s'})^{-1} \\ \times \lim_{1,\ldots,k-1} \prod_{s' \leqslant \sigma_{k-1}} (\lambda'_s - \lambda'_{s'})^{-1} \lim_{k+1,\ldots,m} \prod_{s' > \sigma_k} (\lambda'_s - \lambda'_{s'})^{-1} \\ = -\sum_{k=1}^{p} \lim_{k} \sum_{s=\sigma_{k-1}+1}^{\sigma_k} \left\{ f_k(\lambda'_s, x) \prod_{s'=\sigma_{k-1}+1}^{\sigma_k}{}' (\lambda'_s - \lambda'_{s'})^{-1} \right\}, \end{aligned}$$

where $f_k(\lambda, x) = (\partial/\partial x) F_k(\lambda, x)$. One recognizes the summation over s as the divided difference of order $\nu_k - 1$ of the function $f_k(\lambda, x)$, corresponding to the ν_k 'abscissas' $\lambda'_{\sigma_{k-1}+1}, \ldots, \lambda'_{\sigma_k}$. Hence, by a well-known result, one has for $x > 0$

$$g(x) = -\sum_{k=1}^{p} \frac{1}{(\nu_k - 1)!} \left[\frac{\partial^{\nu_k - 1}}{\partial \lambda^{\nu_k - 1}} f_k(\lambda, x) \right]_{\lambda = \lambda_k}. \tag{2·2}$$

Upon integrating this expression, formula (2·1) results. For $x < 0$, one must substitute $\sum_{p+1}^{m}$ for $-\sum_{1}^{p}$ in (2·2) and $1 - \sum_{p+1}^{m}$ for $\sum_{1}^{p}$ in (2·1).

A particular case of the theorem is

$$P\{\chi^2_{2\nu} > x\} = \frac{1}{(\nu - 1)!} \left[\frac{\partial^{\nu - 1}}{\partial \lambda^{\nu - 1}} \{\lambda^{\nu - 1} e^{-x/(2\lambda)}\} \right]_{\lambda = 1}.$$

Even using this, it is not easily possible to establish in general the equivalence of (2·1) with formula (2·11) of Box (1954).

The formulae (2·1) and (2·2) are very convenient to use when all ν_k are small. With increasing ν_k's, the labour of computing the corresponding derivatives of $F_k(\lambda, x)$, or $f_k(\lambda, x)$ rapidly becomes considerable. The numerical method of the next section is then often easier to apply.

Consider now the case where the variables are non-central. The characteristic function of (1·1) is

$$\phi(t) = \prod_{1}^{m} (1 - 2i\lambda_r t)^{-\frac{1}{2}h_r} \exp\left\{ i \sum_{1}^{m} \frac{\delta_r^2 \lambda_r t}{1 - 2i\lambda_r t} \right\}. \tag{2·3}$$

To obtain the probability density by integration of the inversion formula appears hopeless, except in the particular case $m = 1$. One finds then for the density $p_n(x, \delta^2)$ of $\chi^2_{n;\delta^2}$ the relation

$$p_n(x, \delta^2) = \tfrac{1}{2} \exp\{-\tfrac{1}{2}(x + \delta^2)\} (x^{\frac{1}{2}}/\delta)^{\frac{1}{2}n - 1} I_{\frac{1}{2}n - 1}(\delta x^{\frac{1}{2}}), \tag{2·4}$$

where $I_\nu(z)$ is the Bessel function of purely imaginary argument, of order ν. This formula has been given first by Fisher (1928) who pointed out on the same occasion the relation

$$P\{\chi^2_{2\nu;\delta^2} < x\} = P\{X - Y \geqslant \nu\},$$

derive some simple approximations for $P\{\chi^2_{n;\delta^2} < x\}$. However, expressions in finite terms obtainable for such probabilities, when n is odd, by integrating (2·4) do not seem to have previously been written down explicitly. We shall give the first few of them, which are simple enough to be useful in numerical work. It is necessary to assume $n > 1$. Write then $n = 2m+3$ $(m = 0, 1, \ldots)$. A classical formula permits the expression in finite terms of the Bessel function of half integral order appearing in (2·4). One finds

$$p_n(x, \delta^2) = \frac{1}{2\delta^{n-2}(2\pi)^{\frac{1}{2}}}\left[e^{-\frac{1}{2}a^2}\sum_{r=0}^{m}(-1)^r A_m(r,x) + (-1)^{m+1}e^{-\frac{1}{2}b^2}\sum_{r=0}^{m}A_m(r,x)\right],$$

where
$$a = x^{\frac{1}{2}} - \delta, \quad b = x^{\frac{1}{2}} + \delta, \quad A_m(r,x) = \frac{(m+r)!\,(\delta x^{\frac{1}{2}})^{m-r}}{2^r\, r!(m-r)!}.$$

The above density can be integrated explicitly. If $I(x)$ denotes the cumulative distribution function of a unit normal variable, the result is found to be

$$P\{\chi^2_{n;\delta^2} > x\} = 2 - I(a) - I(b) + \frac{1}{\delta^{n-2}(2\pi)^{\frac{1}{2}}}[K_n(\delta, a) - K_n(\delta, -b)],$$

where $K_n(\delta, x) = \exp\{-\frac{1}{2}x^2\}\, P_n(\delta, x)$ and $P_n(\delta, x)$ is a polynomial in δ of degree $n-3$. The general expression for $P_n(\delta, x)$ is not simple. The polynomials corresponding to the values $n = 3, 5, 7$ and 9 are

$$P_3(\delta, x) \equiv 1, \quad P_5(\delta, x) = 2\delta^2 + x\delta - 1,$$
$$P_7(\delta, x) = 3\delta^4 + 3x\delta^3 + (x^2 - 4)\,\delta^2 - 3x\delta + 3,$$
$$P_9(\delta, x) = 4\delta^6 + 6x\delta^5 + (4x^2 - 10)\,\delta^4 + (x^3 - 15x)\,\delta^3 - (6x^2 - 18)\,\delta^2 + 15x\delta - 15.$$

In the applications, b is often large enough so that one can assume $I(b) = 1$, $K_n(\delta, -b) = 0$.

3. Numerical inversion of the characteristic function

In this section we show that the cumulative distribution function $F(x)$ of the variable Q defined by (1·1) can be obtained quite easily by straightforward numerical integration of an inversion formula. The standard inversion formula gives an expression not for $F(x)$ itself, but for the difference $F(x) - F(0)$. If Q, as is the case, for instance, in (1·2), is not strictly positive, a formula is needed which gives $F(x)$ directly. Such a formula is implicit in work of Gurland (1948) and has been later derived explicitly by Gil-Pelaez (1951), namely

$$F(x) = \tfrac{1}{2} - \frac{1}{\pi}\int_0^\infty t^{-1}\mathcal{J}\{e^{-itx}\phi(t)\}\,dt, \tag{3·1}$$

where $\mathcal{J}(z)$ denotes the imaginary part of z and where the characteristic function $\phi(t)$ of Q is given by (2·3). Using the relations

$$\arg[(1-ibt)^{-v}] = v\tan^{-1}(bt), \quad |(1-ibt)^{-v}| = (1+b^2t^2)^{-\frac{1}{2}v},$$
$$\arg[\exp\{iat/(1-ibt)\}] = at/(1+b^2t^2), \quad |\exp\{iat/(1-ibt)\}| = \exp\{-abt^2/(1+b^2t^2)\},$$

one finds that for the quadratic form Q of (1·1), equation (3·1) can be written, after the substitution $2t = u$ is made,

$$P\{Q > x\} = \tfrac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\sin\theta(u)}{u\rho(u)}\,du, \tag{3·2}$$

where

$$\theta(u) = \frac{1}{2}\sum_1^m [h_r \tan^{-1}(\lambda_r u) + \delta_r^2 \lambda_r u(1+\lambda_r^2 u^2)^{-1}] - \tfrac{1}{2}xu,$$

$$\rho(u) = \prod_1^m (1+\lambda_r^2 u^2)^{\frac{1}{4}h_r} \exp\left\{\frac{1}{2}\sum_1^m (\delta_r \lambda_r u)^2 \Big/ (1+\lambda_r^2 u^2)\right\}.$$

The following relations are immediate

$$\lim_{u\to 0} \frac{\sin\theta(u)}{u\rho(u)} = \frac{1}{2}\sum_1^m \lambda_r(h_r+\delta_r^2) - \tfrac{1}{2}x \tag{3·3}$$

$$\lim_{u\to\infty} \theta(u) = \begin{cases} -\infty & \text{if } x > 0, \\ +\infty & \text{if } x < 0, \\ \frac{\pi}{4}\sum_1^m h_r \lambda_r |\lambda_r|^{-1} & \text{if } x = 0. \end{cases}$$

The function $u\rho(u)$ increases monotonically towards $+\infty$. Therefore, in numerical work the integration in (3·2) will be carried over a finite range $0 \leqslant u \leqslant U$ only; the degree of approximation obtained will depend (apart from rounding-off errors) on two sources of error: (i) the 'error of integration' resulting from the use of an approximate rule for computing

$$I_U = \pi^{-1}\int_0^U [u\rho(u)]^{-1} \sin\theta(u)\, du,$$

and (ii) the 'error of truncation'

$$t_U = \pi^{-1}\int_U^\infty [u\rho(u)]^{-1} \sin\theta(u)\, du.$$

Noticing that $x > y$ implies $x^2(1+x^2)^{-1} > y^2(1+y^2)^{-1}$, one finds that $|t_U|$ can be bounded above by T_U, where

$$T_U^{-1} = \pi k U^k \prod_1^m |\lambda_r|^{\frac{1}{2}h_r} \exp\left\{\frac{1}{2}\sum_1^m \delta_r^2 \lambda_r^2 U^2 (1+\lambda_r^2 U^2)^{-1}\right\},$$

and where we have put $k = \frac{1}{2}\sum_1^m h_r$. One can hopefully expect that T_U will often be satisfactorily small, even for moderate values of U. It does not seem feasible on the other hand to obtain an upper bound for the error of integration resulting from the application of a standard quadrature formula. With equal step formulae (in the application of which the ordinate at the origin is given by (3·3)), a common procedure is then to carry out the integration repeatedly, each time halving the step of integration until two consecutive results agree to the desired accuracy. In order to gather some more detailed information relative to the performance of the two simplest and most commonly used rules, namely, the trapezoidal rule and Simpson's rule, we have preferred to follow a different procedure. Given a tolerance ϵ, numerical integration in (3·2) was carried out by using steps of length $j/10$. The integer j was then increased, one unit at a time, until the largest value was attained such that, for the first even number of steps yielding a bound $T_U \leqslant \epsilon$, the values of the integral obtained by the trapezoidal rule and by Simpson's rule, say J_T and J_S, still satisfy $|J_T - J_S| \leqslant \epsilon$. Results for the values $\epsilon = 0{\cdot}01$ and $\epsilon = 0{\cdot}001$ are given in Table 1. The following unexpected behaviour was observed in all cases: As j increases, J_S departs from the correct value of the integral 'more rapidly' than does J_T. In other words, to achieve a given accuracy, longer steps are permissible with the trapezoidal rule than with Simpson's rule. Some computations made with

a formula (sometimes referred to as Hardy's rule) based on approximation of the integrand with polynomials of degree six seem to indicate that in fact, the higher the degree of the approximation polynomials used, the shorter the step of integration must be. Accordingly, only the results obtained with the trapezoidal rule are recorded in columns (i) and (ii) of Table 1 and are seen to be of much greater accuracy than the value of ϵ would let one

Table 1. *Probability that the quadratic form Q exceeds x*

(i) Integration in (3·3) with the trapezoidal rule and n steps of length j, for tolerance $\epsilon = 0{\cdot}01$; (ii) same for tolerance $\epsilon = 0{\cdot}001$; (iii) exact values; (iv) Pearson's three-moment χ^2 approximation; (v) Patnaik's two-moment χ^2 approximation.

		$\epsilon = 0{\cdot}01$			$\epsilon = 0{\cdot}001$					
Q	x	n	j	(i)	n	j	(ii)	(iii)	(iv)	(v)
$Q_1 = 0{\cdot}6\chi_1^2 + 0{\cdot}3\chi_1^2 + 0{\cdot}1\chi_1^2$	0·1	16	1·9	0·9478	136	1·0	0·9457	0·9458	1·0000	0·9184
	0·7	14	2·2	·5076	114	1·2	·5065	·5064	0·4901	·5079
	2	10	3·0	·1262	86	1·6	·1240	·1240	·1264	·1310
$Q_2 = 0{\cdot}6\chi_2^2 + 0{\cdot}3\chi_2^2 + 0{\cdot}1\chi_2^2$	0·2	8	1·2	0·9903	24	0·8	0·9930	0·9936	1·0000	0·9867
	2	6	1·9	·4020	20	1·0	·3997	·3998	0·3961	·4098
	6	8	1·1	·0159	20	1·0	·0161	·0161	·0162	·0145
$Q_3 = 0{\cdot}6\chi_6^2 + 0{\cdot}3\chi_4^2 + 0{\cdot}1\chi_2^2$	1	6	0·7	0·9980	14	0·4	0·9974	0·9973	0·9984	0·9961
	5	4	1·1	·4354	8	0·7	·4353	·4353	·4346	·4400
	12	8	0·6	·0087	12	0·5	·0088	·0088	·0087	·0080
$Q_4 = 0{\cdot}6\chi_2^2 + 0{\cdot}3\chi_4^2 + 0{\cdot}1\chi_6^2$	1	8	1·1	0·9663	16	0·7	0·9666	0·9666	0·9767	0·9522
	3	6	1·4	·4201	10	1·0	·4195	·4196	·4156	·4330
	8	10	0·8	·0088	14	0·8	·0087	·0087	·0086	·0066
$Q_5 = 0{\cdot}7\chi_{6;6}^2 + 0{\cdot}3\chi_{2;2}^2$	2	6	0·3	0·9938	14	0·2	0·9937	0·9939	0·9928	0·9954
	10	4	0·7	·4098	8	0·4	·4086	·4087	·4089	·4045
	20	6	0·3	·0223	10	0·3	·0221	·0221	·0221	·0230
$Q_6 = 0{\cdot}7\chi_{1;6}^2 + 0{\cdot}3\chi_{1;2}^2$	1	8	0·4	0·9544	48	0·3	0·9549	0·9549	0·9515	0·9719
	6	6	0·7	·4079	36	0·4	·4075	·4076	·4084	·3948
	15	8	0·4	·0224	36	0·4	·0223	·0223	·0223	·0246
$\frac{1}{3}Q_3 + \frac{2}{3}Q_4$	1·5	8	1·2	0·9891	12	0·9	0·9890	0·9891	0·9923	0·9842
	4	4	2·4	·3463*	8	1·4	·3453	·3453	·3449	·3552
	7	8	1·2	·0154	12	1·0	·0154	·0154	0·154	·0131
$\frac{1}{3}Q_3 - \frac{2}{3}Q_4$	−2	6	1·6	0·9103	10	1·2	0·9102	0·9102	0·9011	—
	0	4	2·2	·4052	10	1·3	·4061	·4061	·4221	—
	2·5	8	1·1	·0097	14	0·8	·0098	·0097	·0032	—
$\frac{1}{2}Q_5 + \frac{1}{2}Q_6$	3·5	6	0·6	0·9563	10	0·4	0·9563	0·9563	0·9560	0·9605
	8	4	1·1	·4161	8	0·6	·4152	·4152	·4153	·4101
	13	6	0·6	·0462	8	0·5	·0462	·0462	·0461	·0474
$\frac{1}{2}Q_5 - \frac{1}{2}Q_6$	−2	6	0·6	0·9218	8	0·5	0·9218	0·9218	0·9166†	—
	2	2	3·9	·4685	4	1·4	·4782	·4779	·4810†	—
	7	6	0·6	·0396	10	0·4	·0396	·0396	·0387†	—
$\frac{1}{4}(Q_3 + Q_4 + Q_5 + Q_6)$	3	10	1·0	0·9842	16	0·7	0·9842	0·9842	0·9840	0·9837
	6	6	2·0	·4270*	10	1·2	·4264	·4264	·4264	·4270
	10	12	0·9	·0117	14	0·8	·0117	·0117	·0117	·0116
$\frac{1}{6}(Q_3 - Q_5) + \frac{2}{6}(Q_6 - Q_4)$	−3	12	0·9	0·9861	18	0·7	0·9861	0·9861	0·9904	—
	0	6	2·0	·5180	12	1·1	·5170	·5170	·5154	—
	4	12	0·9	·0152	16	0·8	·0152	·0152	·0139	—

* Corresponds to first maximum of $|J_T - J_S|$. † Wilson–Hilferty approximation.

suspect. We have in several cases computed the correction terms, up to the one involving finite differences of orders one to six, which are to be added to the trapezoidal rule according to Gregory's formula. They have proved to give no reliable estimate of the error of integration resulting from the use of the trapezoidal rule. For $\epsilon = 0{\cdot}01$, it was found in two cases that as j increases, $|J_T - J_S|$ first increases to reach a maximum smaller than 0·01, then decreases again. The results corresponding to the value of j for which $|J_T - J_S|$ attains its first maximum are then recorded. It is clear that under such circumstances, the procedure we have followed is not an adequate one to use in general practice. Another example of this is provided by $Q = \frac{1}{2}Q_5 - \frac{1}{2}Q_6$, for $x = 2$. In this case, it was necessary to increase j up to 3·9 before $|J_T - J_S|$ reached $\epsilon = 0{\cdot}01$, while two steps of length 1·5 already make $T_U < 0{\cdot}01$ and give the better value 0·4787. It is hoped, however, that a sufficient variety of quadratic forms is covered so that Table 1 can serve as a useful guide for the length of the step of integration to be used in any particular case. To help in this purpose, all quadratic forms listed satisfy $\sum_1^m |\lambda_r| = 1$.

The probability density $g(x)$ of the quadratic form Q could be computed by using a formula analogous to (3·2). In fact,

$$g(x) = \pi^{-1} \int_0^\infty [\rho(u)]^{-1} \cos \theta(u)\, du.$$

Due to the absence of the factor u^{-1} in the integrand, numerical integration, for the same accuracy, can be expected to require a slightly larger number of steps than is needed to compute the distribution function.

4. The accuracy of some approximations

Recently Johnson (1959) has shown that Pearson's (1959) three-moment central χ^2 approximation to the distribution of non-central χ^2 is remarkably accurate in both tails of the distribution. It is therefore natural to extend this approximation to the general case of the quadratic form Q of (1·1). Let $E(Q)$ and $\sigma(Q)$ be the mean and standard deviation of Q. Following Pearson, we write for a positive quadratic form Q,

$$Q \cong (\chi^2_{h'} - h')\,(2h')^{-\frac{1}{2}}\,\sigma(Q) + E(Q),$$

and determine h' so that both members have equal third moments. This amounts to taking

$$P\{Q > x\} \cong P\{\chi^2_{h'} > y\}, \tag{4·1}$$

where $\quad h' = c_2^3/c_3^2, \quad y = (x - c_1)\,(h'/c_2)^{\frac{1}{2}} + h', \quad c_j = \sum_1^m \lambda_r^j (h_r + j\delta_r^2) \quad (j = 1, 2, 3).$

If Q is non-positive, the same approximation can be used but one must assume that $c_3 > 0$. Otherwise, approximate the distribution of $-Q$. The values obtained from (4·1) are shown in column (iv) of Table 1. For the quadratic form $\frac{1}{2}Q_5 - \frac{1}{2}Q_6$, one has $h' = 1395{\cdot}85$. In this case, the Wilson–Hilferty approximation was used to evaluate the right-hand member of (4·1). The merits of the three-moment approximation are obvious, particularly in the upper tail of positive quadratic forms. For such forms, the approximate values given by the standard two-moment (Patnaik) approximation are indicated in column (v). It is seen that the three-moment approximation, which requires very little more work, gives a much better fit than is achieved with the Patnaik approximation. While there is an appreciable loss of accuracy for non-positive forms, the approximation still gives useful indications which can suffice for certain practical purposes.

Most of the computations summarized in Table 1 have been performed on the Ferranti 'Mercury' computer at CERN, Geneva. I am much indebted to the late Director General of that organization, Prof. C. J. Bakker, who made it possible for me to use the facilities of the computer service.

I wish to thank the Referee for suggesting some notable improvements in the presentation of the paper.

REFERENCES

BOX, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Ann. Math. Statist.* **25**, 290–302.

DAVENPORT, W. B. & ROOT, W. L. (1958). *An Introduction to the Theory of Random Signals and Noise.* New York: McGraw-Hill.

FISHER, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc.* A, **121**, 654–73.

GIL-PELAEZ, J. (1951). Note on the inversion theorem. *Biometrika*, **38**, 481–2.

GRAD, A. & SOLOMON, H. (1955). Distribution of quadratic forms and some applications. *Ann. Math. Statist.* **26**, 464–77.

GURLAND, J. (1948). Inversion formulae for the distribution of ratios. *Ann. Math. Statist.* **19**, 228–37.

GURLAND, J. (1955). Distribution of definite and indefinite quadratic forms. *Ann. Math. Statist.* **26**, 122–7.

JOHNSON, N. L. (1959). On an extension of the connexion between Poisson and χ^2 distributions. *Biometrika*, **46**, 352–63.

PEARSON, E. S. (1959). Note on an approximation to the distribution of non-central χ^2. *Biometrika*, **46**, 364.

ROE, G. M. & WHITE, G. M. (1961). Probability density functions for correlators with noisy reference signals. *IRE Trans. on Information Theory*, IT-7, 13–18.

SCHEFFÉ, H. (1959). *The Analysis of Variance.* New York: John Wiley and Sons.

Biometrika (1961), **48**, 3 *and* 4, *p.* 427
Printed in Great Britain

Comparative efficiencies of methods of estimating parameters in linear autoregressive schemes

BY K. C. CHANDA
Bombay University

1. INTRODUCTION

It is normal practice to estimate the parameters of a linear autoregressive (a.r.) scheme of a specified order, say p, by minimizing the sum of squares of the residuals. Asymptotically, the resulting estimators are simply related to the first p-serial correlations in the sample and are fully efficient; their properties have been investigated by Mann & Wald (1943).

What remains, however, a problem still unsolved is to determine whether there exist other methods of estimation for which better results are available in short series. In principle one can use information contained in the higher-order serial correlations and it is possible that in series of short length such information will prove useful. This possibility has been suggested by Ghurye (1950), Kendall (1949) and Quenouille (1957). A first step in the study of the properties of the estimators suggested by these authors is to investigate their asymptotic efficiency; this investigation forms the subject of this paper.

2. KENDALL'S METHOD

Let us assume that the a.r. model is

$$\sum_{u=0}^{p} a_u X_{t-u} = Y_t \quad (a_0 = 1),$$

where $\{X_t\}$ is the a.r. process and the $\{Y_t\}$ is a purely random process. We assume, further, that $\mathscr{E}(X_t) = 0$ and $V(Y_t) = 1$. Kendall (1949) has suggested the following normal equations for the estimation of a_u $(u = 1, 2, \ldots, p)$

$$\left(\frac{\partial}{\partial a_u} \sum_{s=1}^{k} \theta_s^2\right)_{\hat{a}_u} = 0 \quad (u = 1, 2, \ldots, p),$$

where

$$\theta_s = \sum_{u=0}^{p} a_u C_{s-u} \quad \text{and} \quad C_r = \sum_{t=1}^{n-|r|} X_t X_{t+|r|}/(n-|r|)$$

and k is an arbitrary positive integer $> p$. Evidently, the equations reduce to

$$\sum_{s=1}^{k} \hat{\theta}_s C_{s-u} = 0 \quad (u = 1, 2, \ldots, p) \tag{1}$$

$(\hat{\theta}_s = \Sigma \hat{a}_u C_{s-u})$ and the estimators are

$$\hat{a}_u = R_{1.u+1}/R_{11} \quad (u = 1, 2, \ldots, p),$$

where

$$\mathbf{R} = \begin{pmatrix} \sum_{s=1}^{k} r_s^2 & \sum_{s=1}^{k} r_s r_{s-1} & \cdots & \sum_{s=1}^{k} r_s r_{s-p} \\ \sum_{s=1}^{k} r_s r_{s-1} & \sum_{s=1}^{k} r_{s-1}^2 & \cdots & \sum_{s=1}^{k} r_{s-1} r_{s-p} \\ \vdots & \vdots & & \vdots \\ \sum_{s=1}^{k} r_s r_{s-p} & \sum_{s=1}^{k} r_{s-1} r_{s-p} & \cdots & \sum_{s=1}^{k} r_{s-p}^2 \end{pmatrix},$$

where $r_s = C_s/C_0$ and R_{ij} = cofactor of the (i,j)th element in $|R|$ $(1 \leqslant i, j \leqslant p+1)$.

If we write $n^{\frac{1}{2}}(\hat{a}_u - a_u) = \delta_u$ it can then be shown that

$$\boldsymbol{\delta}'\mathbf{T}\mathbf{T}' \sim -n^{\frac{1}{2}}\boldsymbol{\theta}'\mathbf{T}', \tag{2}$$

where $\sim$ stands for equivalence in probability and

$$\boldsymbol{\delta}' = (\delta_1, \delta_2, \ldots, \delta_p),$$

$$\mathbf{T} = \begin{pmatrix} \tau_0 & \tau_1 & \cdots & \tau_{k-1} \\ \tau_1 & \tau_0 & \cdots & \tau_{k-2} \\ \vdots & \vdots & & \vdots \\ \tau_{p-1} & \tau_{p-2} & \cdots & \tau_{k-p} \end{pmatrix},$$

$$\boldsymbol{\theta}' = (\theta_1, \theta_2, \ldots, \theta_k),$$

$$\tau_s = \mathscr{E}(C_s).$$

Hence

$$\boldsymbol{\delta}' \sim -n^{\frac{1}{2}}\boldsymbol{\theta}'\mathbf{T}'(\mathbf{T}\mathbf{T}')^{-1}$$

and if $\boldsymbol{\Lambda}$ denotes the asymptotic dispersion matrix of $\boldsymbol{\delta}'$

$$\boldsymbol{\Lambda} = (\mathbf{T}\mathbf{T}')^{-1}\mathbf{T}\mathbf{T}_k\mathbf{T}'(\mathbf{T}\mathbf{T}')^{-1}, \tag{3}$$

where

$$\mathbf{T}_r = \begin{pmatrix} \tau_0 & \tau_1 & \cdots & \tau_{r-1} \\ \tau_1 & \tau_0 & \cdots & \tau_{r-2} \\ \vdots & \vdots & & \vdots \\ \tau_{r-1} & \tau_{r-2} & \cdots & \tau_0 \end{pmatrix} \quad \text{for all } r \geqslant 1.$$

This follows from the fact that $n \operatorname{cov}(\theta_r, \theta_s) \sim \tau_{r-s}$. It is known that when a_u are estimated by the classical method, the asymptotic dispersion matrix $\boldsymbol{\Lambda}_0$ for $n^{\frac{1}{2}}(\hat{a}_u - a_u)$ $(u = 1, 2, \ldots, p)$ is given by

$$\boldsymbol{\Lambda}_0 = \mathbf{T}_p^{-1}. \tag{4}$$

Hence the generalized efficiency of Kendall's method is the ratio

$$e_k = |\mathbf{T}\mathbf{T}'|^2/|\mathbf{T}_p|\,|\mathbf{T}\mathbf{T}_k\mathbf{T}'|. \tag{5}$$

It is not difficult to prove that $e_k \leqslant 1$. As a matter of fact, let us write

$$\boldsymbol{\xi}' = \boldsymbol{\theta}_0'\mathbf{T}_p^{-1},$$

$$\boldsymbol{\eta}' = \boldsymbol{\theta}'\mathbf{T}',$$

where $\boldsymbol{\theta}_0' = (\theta_1, \theta_2, \ldots, \theta_p)$. The dispersion matrix of the vectors $\boldsymbol{\xi}'$ and $\boldsymbol{\eta}'$ will then be

$$\begin{pmatrix} \mathbf{T}_p^{-1} & \mathbf{T}_p^{-1}\mathbf{T}\mathbf{T}' \\ \mathbf{T}\mathbf{T}'\mathbf{T}_p^{-1} & \mathbf{T}\mathbf{T}_k\mathbf{T}' \end{pmatrix}.$$

It follows, therefore, that $\mathbf{T}_p^{-1} - \mathbf{T}_p^{-1}\mathbf{T}\mathbf{T}'(\mathbf{T}\mathbf{T}_k\mathbf{T}')^{-1}\mathbf{T}\mathbf{T}'\mathbf{T}_p^{-1}$ is a semi-positive definite matrix and so

$$|\mathbf{T}_p|^{-1} \geqslant |\mathbf{T}_p|^{-2}|\mathbf{T}\mathbf{T}'|^2|\mathbf{T}\mathbf{T}_k\mathbf{T}'|^{-1},$$

i.e. $e_k \leqslant 1$. Equality holds if $k = p$.

For the particular case $p = 1$, this reduces to

$$e_k = (1-\rho^{2k})^2/\{1+\rho^2-(2k+1)\rho^{2k}+(2k-1)\rho^{2k+2}\} \quad (\rho = -a_1), \tag{6}$$

which can be written in the alternative form

$$e_k = 1-\rho^2(1-\rho^2)\left\{\sum_{r=0}^{k-2}\frac{(r+1)(r+2)}{2}\rho^{2r}+\sum_{r=k-1}^{2k-4}\frac{(2k-r-3)(2k-r-2)}{2}\rho^{2r}\right\}\Big/\sum_{r=0}^{k-1}(2r+1)\rho^{2r}.$$

In particular $e_1 = 1$, $e_2 = 1-\rho^2(1-\rho^2)/(1+3\rho^2) = (1+\rho^2)^2/(1+3\rho^2)$.

3. Ghurye's method

Let us define A_q $(q = 0, \pm 1, \ldots, \pm p)$ by

$$\sum_{u=0}^{p} a_u\xi^u \sum_{u=0}^{p} a_u\xi^{-u} = \sum_{q=-p}^{p} A_q\xi^q \tag{7}$$

and let

$$B_s = \sum_{q=-p}^{p} A_q C_{s-q}.$$

The normal equations for estimating a_u as suggested by Ghurye (1950) are

$$\left(\frac{\partial}{\partial a_u}\sum_{s=1}^{k} B_s^2\right)_{\hat{a}_u} = 0,$$

i.e.

$$\sum_{s=1}^{k}\sum_{q=-p}^{p}\hat{A}_q C_{s-q}\sum_{v=0}^{p}\hat{a}_v C_{s+u-v} = 0 \quad (u = 1, 2, \ldots, p).$$

If we write, as before, $\delta_u = n^{\frac{1}{2}}(\hat{a}_u - a_u)$ it can then be shown after some simplification that for large n

$$\boldsymbol{\delta}'\mathbf{G}\mathbf{G}' \sim -n^{\frac{1}{2}}\boldsymbol{\beta}'\mathbf{G}', \tag{8}$$

where

$$\boldsymbol{\beta}' = (B_1, B_2, \ldots, B_k),$$

$$\mathbf{G} = \begin{pmatrix} g_0 & g_1 & \cdots & g_{k-1} \\ 0 & g_0 & \cdots & g_{k-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & g_{k-p} \end{pmatrix}$$

and

$$\sum_{w=0}^{\infty} g_w\xi^w = \left(\sum_{u=0}^{p} a_u\xi^u\right)^{-1} \quad (g_0 = 1).$$

It has, further, been proved by Bartlett & Diananda (1950) that $n^{\frac{1}{2}}B_1, n^{\frac{1}{2}}B_2, \ldots, n^{\frac{1}{2}}B_k$ have, asymptotically, independent normal distributions each with zero mean and unit standard deviation. Hence,

$$\boldsymbol{\delta}' \sim -n^{\frac{1}{2}}\boldsymbol{\beta}'\mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}$$

and if the asymptotic dispersion matrix of $\boldsymbol{\delta}'$ is $\boldsymbol{\Lambda}$,

$$\boldsymbol{\Lambda} = (\mathbf{G}\mathbf{G}')^{-1}\mathbf{G}\mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1} = (\mathbf{G}\mathbf{G}')^{-1}. \tag{9}$$

Hence the generalized efficiency is the ratio

$$e_k = |\mathbf{G}\mathbf{G}'|/|\mathbf{T}_p|. \tag{10}$$

Here again, if we consider the q.f. $\mathbf{y}'(\mathbf{T}_p - \mathbf{GG}')\mathbf{y}$, where $\mathbf{y}' = (y_1, y_2, \ldots, y_p)$ is an arbitrary vector, we can easily show that the q.f.

$$\sum_{s=k+1}^{\infty} \mathbf{y}'\boldsymbol{\gamma}_s\boldsymbol{\gamma}_s'\mathbf{y} \geqslant 0, \quad \text{where} \quad \boldsymbol{\gamma}_s' = (g_{s-1}, g_{s-2}, \ldots, g_{s-p}).$$

Hence $\mathbf{T}_p - \mathbf{GG}'$ is semi-positive definite and $|\mathbf{T}_p| \geqslant |\mathbf{GG}'|$, i.e. $e_k \leqslant 1$.

For the particular case $p = 1$,

$$e_k = 1 - \rho^{2k} \quad (\rho = -a_1). \tag{11}$$

Thus both Kendall's and Ghurye's methods lead to asymptotically inefficient estimators.

4. Quenouille's method

Quenouille (1957) has suggested a different method. It leads to efficient estimators and at the same time it is different from the classical method, in so far as it is an attempt to remove the bias of order n^{-1} in the least squares estimators, but it must be admitted the normal equations are difficult to solve.

As Quenouille pointed out, the methods of estimation so far employed do not make use of the fact that the functions $\theta_1, \theta_2, \ldots, \theta_k$ are, asymptotically, related and that the dispersion matrix itself will depend on the parameters of the a.r. process. One, may, therefore, consider the asymptotic likelihood function for a_u based on θ_s's (Quenouille has used the serial correlations r in the definition of θ_s's to include the situations where $V(Y_t)$ is unknown). The normal equations are then derived by maximizing the logarithm of this likelihood function and take the form

$$\hat{\tau}_0^2\left\{\sum_{s,t=1}^{k} \theta_s\theta_t\left(\frac{\partial\tau^{st}}{\partial a_u}\right)_{\hat{a}_u} + 2\sum_{s,t=1} \theta_s\hat{\tau}^{st}r_{t-u}\right\} + 2\hat{\tau}_0\left(\frac{\partial\tau_0}{\partial a_u}\right)_{\hat{a}_u}\sum_{s,t=1}^{k}\theta_s\theta_t\hat{\tau}^{st}$$
$$- 2k/n\left(\frac{\partial\tau_0}{\partial a_u}\right)_{\hat{a}_u}\hat{\tau}_0^{-1} + 1/n\left(\frac{\partial|\mathbf{T}_k|}{\partial a_u}\right)_{\hat{a}_u}\Big/|\hat{\mathbf{T}}_k| = 0 \quad (u = 1, 2, \ldots, p), \tag{12}$$

where $((\tau^{st})) = \mathbf{T}_k^{-1}$ and $\hat{\tau}^{st}$ is the same as τ^{st} except that a_u have been replaced by $\hat{a}_u$. When terms of order $n^{-\frac{1}{2}}$ only are retained it can easily be seen that the left-hand side expression of (12) reduces to a multiple of $\hat{\theta}_u$ and hence (12) represent the classical least-squares equations. On the other hand, if we retain terms of order n^{-1} as well, this may help towards correction for bias of order n^{-1}.

For the particular case $p = 1$ and $k = 2$ a slight modification in Quenouille's estimator will lead to

$$\hat{a}_1 = -r_1 - 2r_1(r_2 - r_1^2)^2/(1 - r_1^2)^2. \tag{13}$$

The estimator is, asymptotically, efficient and its bias to order n^{-1} is zero. When $p = 2$ and $k = 3$, the equations (12) reduce to

$$\left.\begin{aligned}
&\hat{\tau}_0^{-1}[\theta_1\theta_2 + \theta_2\theta_3 + \hat{a}_1\theta_2^2 + \theta_1(1 + \hat{a}_1r_1 + \hat{a}_2r_2) + \theta_2\{\hat{a}_1(1 + r_2) + r_1(1 + \hat{a}_1^2 - \hat{a}_2^2)\}\\
&\quad + \theta_3(r_2 + \hat{a}_1r_1 + \hat{a}_2)] - 2\hat{\rho}_1(1 - \hat{a}_2)[\theta_1^2 + \theta_3^2 + 2\hat{a}_1(\theta_1\theta_2 + \theta_2\theta_3) + 2\hat{a}_2\theta_1\theta_3\\
&\quad + \theta_2^2(1 + \hat{a}_1^2 - \hat{a}_2^2)] + 5\hat{\tau}_0^{-2}\hat{\rho}_1(1 - \hat{a}_2)/n = 0,\\
&\hat{\tau}_0^{-1}[\theta_1\theta_3 - \hat{a}_2\theta_2^2 + (\theta_1 + \theta_3)\{(1 + \hat{a}_2)r_1 + \hat{a}_1\} + \theta_2(1 + 2\hat{a}_1r_1 + \hat{a}_1^2 - \hat{a}_2^2)]\\
&\quad - 2(\hat{\rho}_1^2 - \hat{a}_2)[\theta_1^2 + \theta_3^2 + 2\hat{a}_1(\theta_1\theta_2 + \theta_2\theta_3) + 2\hat{a}_2\theta_1\theta_3 + \theta_2^2(1 + \hat{a}_1^2 - \hat{a}_2^2)]\\
&\quad + 6\hat{\tau}_0^{-2}(\hat{\rho}_1^2 - \hat{a}_2)/n - \hat{\tau}_0^{-2}(\hat{\rho}_2 - \hat{a}_2)/n = 0.
\end{aligned}\right\} \tag{14}$$

Solutions to the above equations are, no doubt, difficult to obtain. It can, however, be seen that the classical least squares estimators $\hat{a}_1$, $\hat{a}_2$ will provide first approximations to the above solutions and better approximations can be obtained by adding the terms

$$2\hat{\theta}_3^2 r_1 \hat{\tau}_0^2 - (6r_1 + \hat{a}_1)/n$$

and

$$-2\hat{\theta}_3^2 \hat{a}_2 \hat{\tau}_0^2 + 4\hat{a}_2/n$$

to $\hat{a}_1$ and $\hat{a}_2$, respectively, $\hat{\theta}_3$ being obtained by replacing the parameters a_1, a_2 in θ_3 by their estimates $\hat{a}_1$, $\hat{a}_2$ and $\hat{\tau}_0 = 1/(1-r_1^2)(1-\hat{a}_2^2)$.

One reason for introducing such a variety of estimators is to examine with what precision they estimate the higher-order auto-correlations. It is known that for an a.r. process of order p, $\Sigma a_u \tau_{s-u} = 0$ for $s \geqslant 1$. Consequently the estimators $\hat{\theta}_s = \sum_{u=0}^{p} \hat{a}_u C_{s-u}$ should have as small a dispersion determinant as possible for as many values of s as ought to be considered $(p+1 \leqslant s \leqslant k)$. In other words if Λ is the asymptotic dispersion matrix of $n^{\frac{1}{2}}\hat{\theta}_{p+1}$, $n^{\frac{1}{2}}\hat{\theta}_{p+2}$, ..., $n^{\frac{1}{2}}\hat{\theta}_k$, then the relative efficiencies of the different methods we have discussed above should be measured by the ratios of the corresponding determinants $|\mathbf{\Lambda}|$. Considered from this point view we have for the classical least squares method (as well as Quenouille's method)

$$\mathbf{\Lambda}_0 = \mathbf{T}_{k-p} - \mathbf{T}_0' \mathbf{T}_p^{-1} \mathbf{T}_0,$$

where

$$\mathbf{T}_0 = \begin{pmatrix} \tau_p & \tau_{p+1} & \dots & \tau_{k-1} \\ \tau_{p-1} & \tau_p & \dots & \tau_{k-2} \\ \vdots & \vdots & & \vdots \\ \tau_1 & \tau_2 & \dots & \tau_{k-p} \end{pmatrix}. \tag{15}$$

For Kendall's method

$$\mathbf{\Lambda}_K = \text{cofactor of the } p\text{th order principal minor in } [\mathbf{I} - \mathbf{T}'(\mathbf{T}\mathbf{T}')^{-1}\mathbf{T}]\,\mathbf{T}_k[\mathbf{I} - \mathbf{T}'(\mathbf{T}\mathbf{T}')^{-1}\mathbf{T}] \tag{16}$$

and for Ghurye's method

$$\mathbf{\Lambda}_G = \mathbf{T}_{k-p} + \mathbf{T}_0'(\mathbf{G}\mathbf{G}')^{-1}\mathbf{T}_0 - \mathbf{G}_0\mathbf{G}'(\mathbf{G}\mathbf{G}')^{-1}\mathbf{T}_0 - \mathbf{T}_0'(\mathbf{G}\mathbf{G}')^{-1}\mathbf{G}\mathbf{G}_0', \tag{17}$$

where

$$\mathbf{G}_0 = \begin{pmatrix} 0 & \dots & 0 & g_0 & \dots & g_{k-p-1} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & g_0 \end{pmatrix}.$$

It may be noted that if we use serial correlations all throughout instead of the serial covariances, the dispersion matrices will in each case be multiplied by τ_0^{-2}.

For the particular case $p = 1$

$$\mathbf{\Lambda}_0 = ((\lambda_{rs,0})),$$

$$\mathbf{\Lambda}_K = ((\lambda_{rs,K})),$$

$$\mathbf{\Lambda}_G = ((\lambda_{rs,G})),$$

where $\lambda_{rs,0} = (\rho^{|r-s|} - \rho^{r+s})/(1-\rho^2)$,

$$\lambda_{rs,K} = [\rho^{|r-s|} - \rho^{r+s}\{(r+s)(1-\rho^2) + 2 - \rho^{2k-2r} - \rho^{2k-2s}\}/(1-\rho^{2k}) + \rho^{r+s}\{1+\rho^2-(2k+1)\rho^{2k}+(2k-1)\rho^{2k+2}\}/(1-\rho^{2k})^2]/(1-\rho^2).$$

$$\lambda_{rs,G} = [\rho^{|r-s|} - \rho^{r+s}(1-\rho^{2k-2r}-\rho^{2k-2s})/(1-\rho^{2k})]/(1-\rho^2) \quad (\rho = -a_1).$$

For $k = 2$

$$\Lambda_0 = 1,$$

$$\Lambda_K = 1/(1+\rho^2)^2,$$

$$\Lambda_G = (1-\rho^2+\rho^4)/\{(1-\rho^2)(1-\rho^4)\}.$$

It, of course, remains to consider the terms of order n^{-1} in the expectations and variances of the various standardized estimators $n^{\frac{1}{2}}(\hat{a}_u - a_u)$ we have considered above. This will give us some rough idea of the comparative efficiencies of these estimators in small samples. The author intends to present the results of this investigation in some future publications.

REFERENCES

BARTLETT, M. S. & DIANANDA, P. H. (1950). Extensions of Quenouille's test for autoregressive schemes. *J. R. Statist. Soc.* B, **12**, 108–15.

GHURYE, S. G. (1950). A method of estimating the parameters of an autoregressive time-series. *Biometrika*, **37**, 173–8.

KENDALL, M. G. (1949). The estimation of parameters in linear autoregressive time-series. *Econometrica*, **17** (Suppl.), 44–57.

MANN, H. B. & WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica*, **11**, 173–220.

QUENOUILLE, M. H. (1957). *The Analysis of Multiple Time-Series*. London: Charles Griffin and Co.

Biometrika (1961), **48**, 3 *and* 4, *p.* 433
With 3 *text figures*
Printed in Great Britain

Tables of the Freeman–Tukey transformations for the binomial and Poisson distributions*

BY FREDERICK MOSTELLER AND CLEO YOUTZ
Harvard University

SUMMARY

We present a table of the Freeman–Tukey variance stabilizing arc-sine transformation for the binomial distribution together with properties of the transformation. Entries in the table are

$$\theta = \frac{1}{2}\left\{\arcsin\sqrt{\left(\frac{x}{n+1}\right)} + \arcsin\sqrt{\left(\frac{x+1}{n+1}\right)}\right\},$$

where n is the sample size and x is the number of successes observed in a binomial experiment. Values of θ are given in degrees, to two decimal places, for $n = 1\,[1]\,50$ and $x = 0\,[1]\,n$.

In addition, for completeness, we give a table of the corresponding square-root transformation to two decimal places for use with Poisson counts. The observed count is x ($x = 0\,[1]\,50$) and the transformed values are

$$g = \sqrt{x} + \sqrt{(x+1)};$$

the squares of the transformed values are also given for use in analysis of variance computations.

1. INTRODUCTION

Transformations are often used in the analysis of data to improve linearity of regression, to improve normality of distribution, and to stabilize approximately the variance, when it might otherwise depend strongly upon a parameter. Freeman & Tukey (1950) introduced the transformations tabled here to stabilize the variance of binomial and of Poisson counts. (A good account of variance-stabilizing transformations is given by Eisenhart (1947).) The table may also be helpful in using methods developed by Gupta & Sobel (1958) to select a subset of populations better than a standard.

In using the Freeman–Tukey transformations, we found need for tables to speed the work, and they are presented here. The Freeman–Tukey transformation for the binomial number of successes x observed in n independent trials is the averaged angular transformation

$$\theta = \frac{1}{2}\left\{\arcsin\sqrt{\left(\frac{x}{n+1}\right)} + \arcsin\sqrt{\left(\frac{x+1}{n+1}\right)}\right\}. \tag{1}$$

Table 5 gives values of θ in degrees to two decimals for $n = 1\,[1]\,50$ and $x = 0\,[1]\,n$. When θ is measured in degrees, it has variance σ_θ^2 tending to the asymptotic variance

$$\sigma_{\theta\infty}^2 = \frac{821}{n+\frac{1}{2}} \quad \text{(squared degrees)}, \tag{2}$$

for a substantial range of p if n is not too small.

* This work was facilitated by a grant from The Ford Foundation and by the Laboratory of Social Relations, Harvard University.

We tabled the function in degrees rather than radians because we wanted to be able to follow from our table into the arc-sine table by Bliss (1946) for n's larger than 50. We did not consider making the correspondence with the arc-sine table by Stevens (1953) because we had already computed our table before we were aware of his.

For Poisson counts, the corresponding transformation is

$$g = \sqrt{x} + \sqrt{(x+1)}, \tag{3}$$

with variance approximately 1, provided the Poisson parameter exceeds 1. Table 6 gives values of g and of g^2 for $x = 0\,[1]\,50$. We found that the table of g^2 saved a little time in desk calculations.

2. SOME PROPERTIES OF THE TRANSFORMATIONS

The arc-sine transformation. For selected numbers of trials, n, Table 1 shows, approximately, the value of the probability of success on a single trial p (and of $1-p$) at which the maximum of σ_θ^2 occurs, the values of σ_θ^2 at the maximum and of $\sigma_{\theta\infty}^2$, and the ratio $\sigma_\theta^2/\sigma_{\theta\infty}^2$. When σ_θ^2 is plotted against p, as n increases, the positions of the maximum variance move toward the extremes, $p = 0$ and $p = 1$.

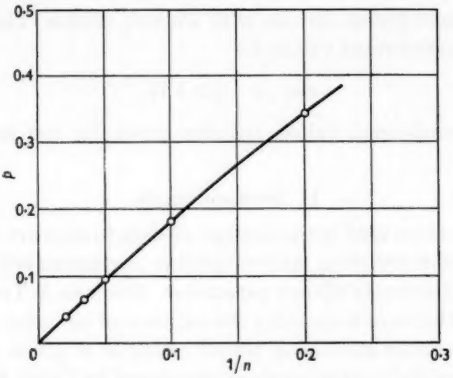

Fig. 1. Values of p ($< \frac{1}{2}$) for maximum σ_θ^2 plotted against $1/n$, $n \geqslant 5$.

Table 1. *Value of p that maximizes σ_θ^2 for the arc-sine transformation, maximum σ_θ^2, $\sigma_{\theta\infty}^2$, and maximum $\sigma_\theta^2/\sigma_{\theta\infty}^2$*

Sample size n	Maximizing values of p	Maximizing values of $1-p$	Max σ_θ^2	$\sigma_{\theta\infty}^2 = \frac{821}{n+\frac{1}{2}}$	Max $\frac{\sigma_\theta^2}{\sigma_{\theta\infty}^2}$
1	0·5	—	506·25	547·33	0·925
2	0·5	—	374·50	328·40	1·140
3	0·5	—	267·19	234·57	1·139
5	0·344	0·656	157·42	149·27	1·055
10	0·181	0·819	82·08	78·19	1·050
20	0·097	0·903	42·20	40·05	1·054
30	0·066	0·934	28·40	26·92	1·055
50	0·040	0·960	17·17	16·26	1·056
∞	Poisson		—	—	1·061

The value of p that maximizes the variance is approximately a linear function of $1/n$ for large n, see Fig. 1. This is explained below in the discussion of the square-root transformation for the Poisson.

Fig. 2 shows $\sigma_\theta^2/\sigma_{\theta\infty}^2$ plotted against p for $n = 1, 3, 5, 10$. At $n = 5$, the curve is beginning to assume its characteristic shape, though the ears (the shapes near the left and right maxima) are quite flat as yet. For $n = 5$, the nearly flat portion in the interval $0{\cdot}3 \leqslant p \leqslant 0{\cdot}7$ appears to have a relative minimum at $p = \frac{1}{2}$.

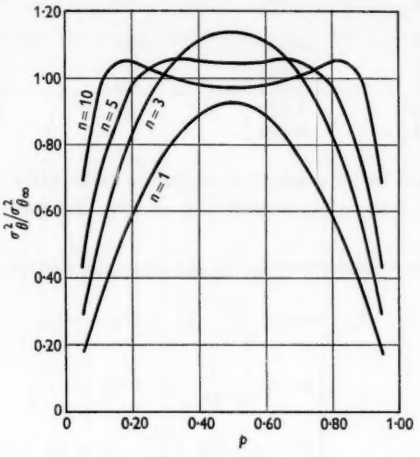

Fig. 2. Plot of $\sigma_\theta^2/\sigma_{\theta\infty}^2$ against p for $n = 1, 3, 5, 10$.

Table 2. *Values of* σ_θ^2 *and of* $\sigma_\theta^2/\sigma_{\theta\infty}^2$ *for* $p = \frac{1}{2}$

Sample size n	σ_θ^2	$\dfrac{\sigma_\theta^2}{\sigma_{\theta\infty}^2}$
1	506·25	0·925
2	374·50	1·140
3	267·19	1·139
4	198·98	1·091
5	156·02	1·045
10	76·05	0·973
20	39·20	0·979
30	26·51	0·985
50	16·10	0·990

One might suppose from the plots of variances of other arc-sine transformations, such as given by Eisenhart (1947) and from those of Fig. 2, that a relative minimum for the variance would occur at $p = \frac{1}{2}$ for large values of n. Our impression from calculations is that the Freeman–Tukey transformation has a relative maximum at $p = \frac{1}{2}$ for $n = 20$ and $n = 50$, where this question was investigated, though the curve of variances is quite flat for a long interval about $p = \frac{1}{2}$. Table 2 gives values of σ_θ^2 at $p = \frac{1}{2}$ for various values of n.

Table 3 shows both σ_θ^2 and $\sigma_{\theta\infty}^2$ for $n = 50$ and for various values of p. The values of p are chosen to illustrate the behaviour to the left and right of the high maximum near

Table 3. σ_θ^2 *and* $\sigma_\theta^2/\sigma_{\theta\infty}^2$ *for* $n = 50$ *for various values of* p

p	σ_θ^2	$\sigma_\theta^2/\sigma_{\theta\infty}^2$	p	σ_θ^2	$\sigma_\theta^2/\sigma_{\theta\infty}^2$
0·01	10·94	0·673	0·10	16·15	0·994
·02	15·34	0·943	·11	16·09	·990
·03	16·85	1·037	·12	16·06	·988
			·13	16·04	·986
0·039	17·17	1·056	·14	16·03	·986
·040	17·17	1·056	·15	16·02	·985
·041	17·17	1·056	·16	16·02	·985
0·05	17·05	1·049	0·20	16·04	0·986
·06	16·81	1·034	·30	16·07	·989
·07	16·57	1·019	·40	16·09	·990
·08	16·38	1·008	·50	16·10	·990
·09	16·24	0·999			

$p = 0{\cdot}040$. There appears to be a relative minimum between $p = 0{\cdot}15$ and $p = 0{\cdot}20$. In the interval $0{\cdot}07 \leqslant p \leqslant 0{\cdot}93$, σ_θ^2 is within 2 % of $\sigma_{\theta\infty}^2$. In the interval $0{\cdot}02 \leqslant p \leqslant 0{\cdot}98$, σ_θ^2 is within 6 % of $\sigma_{\theta\infty}^2$.

An alternative arc-sine transformation for the binomial distribution is

$$\theta' = \arcsin\sqrt{\frac{x}{n}} \quad (1 \leqslant x \leqslant n-1),$$

$$\theta' = \begin{cases} \arcsin\sqrt{\dfrac{1}{4n}} & (x = 0), \\ 90° - \arcsin\sqrt{\dfrac{1}{4n}} & (x = n). \end{cases}$$

For $n = 50$, Fig. 3 relates $\sigma_\theta^2/\sigma_{\theta\infty}^2$ to p and $\theta_{\theta'}^2/\sigma_{\theta'\infty}^2$ to p (where $\sigma_{\theta'\infty}^2 = 821/n$). Only the left half of the curve is plotted ($0 \leqslant p \leqslant 0{\cdot}5$). The variance of the Freeman–Tukey transformation is flatter over a longer region, as of course it is intended to be. Table 4 gives some numerical values of the ratios $\sigma_{\theta'}^2/\sigma_{\theta'\infty}^2$ for $n = 50$.

Table 4. $\sigma_{\theta'}^2$ *and* $\sigma_{\theta'}^2/\sigma_{\theta'\infty}^2$ *for* $n = 50$ *for various values of* p

p	$\sigma_{\theta'}^2$	$\sigma_{\theta'}^2/\sigma_{\theta'\infty}^2$
0·1	17·91	1·091
·2	17·11	1·042
·3	16·87	1·027
·4	16·78	1·022
·5	16·76	1·020

In Fig. 3, the reader will observe that over the long flat region the variance ratio curve for θ falls below 1. If one wished a better match between σ_θ^2 and $\sigma_{\theta\infty}^2$ over this interval, he might use $\sigma_{\theta\infty}^2 = 821/(n+1)$ rather than $821/(n+\frac{1}{2})$ for $10 \leqslant n \leqslant 50$. If he does, then a larger percentage error is committed for p's in the neighbourhood of the ears.

The square-root transformation. Table 6 gives the square-root transformation for the Poisson distribution. If λ is the parameter (mean or expectation) of the Poisson, then for $\lambda = 0$, $\sigma_y^2 = 0$, and as λ increases, σ_y^2 increases to a maximum of about 1·061 in the neighbourhood of $\lambda = 2{\cdot}1$. Thereafter σ_y^2 decreases, apparently never falling below 1. A graph of σ_y^2 is given by Freeman & Tukey (1950). For $\lambda \geqslant 1$, σ_y^2 is within 6 % of its asymptotic value 1.

When a binomial p is small and n is large, the binomial distribution is approximated by the Poisson with parameter $\lambda = np$. Furthermore, when θ is measured in radians, for small values of $\overline{p}$, we have $\arcsin\sqrt{\overline{p}} \approx \sqrt{\overline{p}}$. Thus the square-root transformation for the Poisson and the arc-sine transformation are essentially equivalent in distribution provided that p and $(x+1)/(n+1)$ are small. We can expect therefore, even for large values of n, that the ears of plots of the variance against p for the arc-sine transformation do not vanish, but that the maximum value of $\sigma^2_\theta/\sigma^2_{\theta\infty}$ is about 6 %, and that this maximum occurs near the value of p where $np = 2{\cdot}1$. For $n = 50$, we found the maximizing value of p to be 0·040 (Table 1), as compared with the Poisson estimate 0·042. This explains the approximately linear relation shown in Fig. 1 between the maximizing value of p and $1/n$.

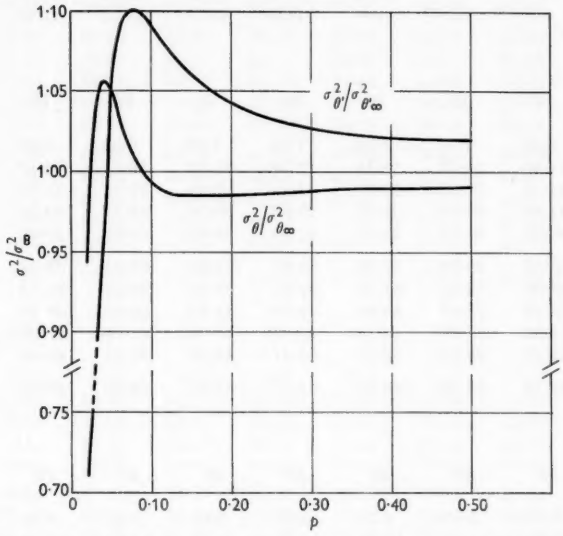

Fig. 3. Graph relating $\sigma^2_\theta/\sigma^2_{\theta\infty}$ to p and relating $\sigma^2_{\theta'}/\sigma^2_{\theta'\infty}$ to p for $n = 50$.

The behaviour of the ears (their persistence) with increasing values of n is reminiscent of Gibbs's phenomenon when a Fourier series is fitted to a discontinuous function. Perhaps the analogy is not far-fetched since the transformation is designed to fit a continuous function with ordinates $y(p) = \sigma^2_\theta/\sigma^2_{\theta\infty}$ and end-points $(0, 0)$ and $(1, 0)$ to the function $f(p) = 1$ $(0 < p < 1)$.

References and tables used in calculations

Bliss, C. I. (1946). Angles corresponding to percentages, Angle = Arc sin $\sqrt{}$(percentage), from *Statistical Methods*, 4th ed., pp. 449–50 by G. W. Snedecor. Ames, Iowa: The Iowa State College Press.

Eisenhart, C. (1947). Inverse sine transformation of proportions, from *Selected Techniques of Statistical Analysis*, by Statistical Research Group, Columbia University, with C. Eisenhart, M. W. Hastay, and W. A. Wallis (eds.), pp. 395–416. New York: McGraw-Hill Book Co., Inc.

Freeman, M. F. & Tukey, J. W. (1950). Transformations related to the angular and the square root. *Ann. Math. Statist.* **21**, 607–11.

Table 5. *Table of Freeman–Tukey arc-sine transformation for binomial proportions**

The table entries are $\theta = \frac{1}{2}\left\{\text{arc sin}\sqrt{\left(\frac{x}{n+1}\right)} + \text{arc sin}\sqrt{\left(\frac{x+1}{n+1}\right)}\right\}$,

where x = number of successes observed, n = sample size.

x \ n	1	2	3	4	5	6	7	8	9	10
0	22·50	17·63	15·00	13·28	12·05	11·10	10·35	9·74	9·22	8·77
1	*67·50*	45·00	37·50	32·90	29·68	27·26	25·35	23·80	22·50	21·39
2	—	*72·37*	52·50	45·00	40·13	36·60	33·88	31·69	29·89	28·36
3	—	—	*75·00*	*57·10*	*49·87*	45·00	41·38	38·54	36·22	34·28
4	—	—	—	*76·72*	*60·32*	*53·40*	*48·62*	45·00	42·12	39·74
5	—	—	—	—	*77·95*	*62·74*	*56·12*	*51·46*	*47·88*	45·00

x \ n	11	12	13	14	15	16	17	18	19	20
0	8·39	8·05	7·75	7·48	7·24	7·02	6·82	6·63	6·46	6·30
1	20·44	19·60	18·85	18·19	17·59	17·05	16·55	16·10	15·68	15·29
2	27·05	25·90	24·89	23·99	23·18	22·45	21·78	21·17	20·61	20·09
3	32·63	31·20	29·94	28·83	27·83	26·93	26·11	25·36	24·68	24·04
4	37·73	36·01	34·51	33·18	31·99	30·93	29·97	29·09	28·28	27·54
5	42·60	40·56	38·80	37·25	35·87	34·64	33·54	32·53	31·61	30·76
6	*47·40*	45·00	42·95	41·16	39·59	38·18	36·92	35·78	34·74	33·79
7	*52·27*	*49·44*	*47·05*	45·00	43·20	41·62	40·20	38·91	37·75	36·69
8	*57·37*	*53·99*	*51·20*	*48·84*	*46·80*	45·00	43·41	41·97	40·68	39·50
9	*62·95*	*58·80*	*55·49*	*52·75*	*50·41*	*48·38*	*46·59*	45·00	43·57	42·26
10	*69·56*	*64·10*	*60·06*	*56·82*	*54·13*	*51·82*	*49·80*	*48·03*	*46·43*	45·00

x \ n	21	22	23	24	25	26	27	28	29	30
0	6·15	6·02	5·89	5·77	5·65	5·55	5·45	5·35	5·26	5·17
1	14·93	14·59	14·28	13·98	13·71	13·44	13·20	12·96	12·74	12·53
2	19·61	19·16	18·74	18·35	17·98	17·63	17·30	16·99	16·70	16·42
3	23·46	22·91	22·40	21·92	21·48	21·05	20·66	20·28	19·93	19·59
4	26·86	26·22	25·63	25·07	24·55	24·06	23·60	23·17	22·76	22·37
5	29·98	29·25	28·58	27·95	27·36	26·81	26·29	25·79	25·33	24·89
6	32·91	32·10	31·34	30·64	29·98	29·37	28·79	28·24	27·72	27·24
7	35·71	34·81	33·98	33·20	32·47	31·79	31·16	30·55	29·99	29·45
8	38·42	37·43	36·51	35·66	34·86	34·12	33·42	32·77	32·15	31·57
9	41·08	39·99	38·98	38·05	37·18	36·38	35·62	34·91	34·24	33·61
10	43·70	42·50	41·41	40·39	39·45	38·58	37·76	36·99	36·27	35·58
11	*46·30*	45·00	43·80	42·70	41·68	40·74	39·85	39·03	38·25	37·52
12	*48·92*	*47·50*	*46·20*	45·00	43·90	42·87	41·92	41·03	40·20	39·42
13	*51·58*	*50·01*	*48·59*	*47·30*	*46·10*	45·00	43·98	43·02	42·13	41·29
14	*54·29*	*52·57*	*51·02*	*49·61*	*48·32*	*47·13*	*46·02*	45·00	44·04	43·15
15	*57·09*	*55·19*	*53·49*	*51·95*	*50·55*	*49·26*	*48·08*	*46·98*	*45·96*	45·00

* For an entry x not in the table ($n \leqslant 50$), take the complement with respect to 90° of the entry for $n-x$, e.g. for $n = 16$, $x = 11$, take $90 - 34·64 = 55·36$. The entries printed in italics are in fact, complements of entries printed higher in that column.

Table 5 (*cont.*)

x \ n	31	32	33	34	35	36	37	38	39	40
0	5·09	5·01	4·94	4·87	4·80	4·73	4·67	4·61	4·55	4·49
1	12·33	12·14	11·96	11·78	11·61	11·45	11·30	11·15	11·01	10·87
2	16·15	15·90	15·66	15·43	15·21	14·99	14·79	14·60	14·41	14·23
3	19·27	18·96	18·67	18·39	18·12	17·87	17·63	17·39	17·16	16·95
4	21·99	21·64	21·30	20·98	20·68	20·38	20·10	19·83	19·57	19·32
5	24·47	24·07	23·69	23·33	22·99	22·66	22·34	22·04	21·75	21·47
6	26·77	26·33	25·91	25·51	25·13	24·76	24·41	24·08	23·76	23·45
7	28·94	28·46	28·00	27·56	27·15	26·75	26·36	26·00	25·65	25·31
8	31·01	30·49	29·99	29·52	29·06	28·63	28·22	27·82	27·44	27·08
9	33·01	32·44	31·90	31·39	30·90	30·44	29·99	29·57	29·16	28·77
10	34·94	34·33	33·75	33·20	32·68	32·18	31·71	31·25	30·81	30·40
11	36·83	36·18	35·56	34·97	34·41	33·88	33·37	32·88	32·42	31·97
12	38·68	37·98	37·32	36·70	36·10	35·53	34·99	34·48	33·98	33·51
13	40·50	39·76	39·06	38·39	37·76	37·16	36·58	36·04	35·51	35·01
14	42·31	41·52	40·77	40·06	39·39	38·75	38·15	37·57	37·02	36·49
15	44·10	43·26	42·47	41·72	41·01	40·33	39·69	39·08	38·50	37·94
16	*45·90*	45·00	44·16	43·36	42·61	41·90	41·22	40·57	39·96	39·37
17	*47·69*	*46·74*	*45·84*	45·00	44·20	43·45	42·74	42·06	41·41	40·79
18	*49·50*	*48·48*	*47·53*	*46·64*	*45·80*	45·00	44·25	43·53	42·85	42·20
19	*51·32*	*50·24*	*49·23*	*48·28*	*47·39*	*46·55*	*45·75*	45·00	44·28	43·60
20	*53·17*	*52·02*	*50·94*	*49·94*	*48·99*	*48·10*	*47·26*	*46·47*	*45·72*	45·00

x \ n	41	42	43	44	45	46	47	48	49	50
0	4·44	4·39	4·34	4·29	4·24	4·19	4·15	4·11	4·07	4·02
1	10·74	10·61	10·49	10·37	10·26	10·15	10·04	9·93	9·83	9·74
2	14·05	13·89	13·72	13·57	13·42	13·27	13·13	12·99	12·86	12·73
3	16·74	16·54	16·34	16·15	15·97	15·80	15·63	15·46	15·30	15·15
4	19·08	18·85	18·62	18·41	18·20	18·00	17·80	17·62	17·43	17·26
5	21·20	20·94	20·69	20·44	20·21	19·99	19·77	19·56	19·35	19·15
6	23·15	22·86	22·59	22·32	22·07	21·82	21·58	21·35	21·12	20·90
7	24·99	24·67	24·37	24·08	23·80	23·53	23·27	23·02	22·78	22·54
8	26·73	26·39	26·06	25·75	25·45	25·16	24·88	24·60	24·34	24·09
9	28·39	28·03	27·68	27·35	27·02	26·71	26·41	26·12	25·83	25·56
10	29·99	29·61	29·24	28·88	28·53	28·20	27·88	27·57	27·27	26·98
11	31·55	31·14	30·74	30·36	29·99	29·64	29·30	28·97	28·65	28·34
12	33·06	32·62	32·20	31·80	31·41	31·04	30·68	30·33	30·00	29·67
13	34·53	34·07	33·63	33·21	32·80	32·40	32·02	31·66	31·30	30·96
14	35·98	35·50	35·03	34·58	34·15	33·74	33·34	32·95	32·58	32·22
15	37·41	36·90	36·41	35·93	35·48	35·05	34·63	34·22	33·83	33·45
16	38·81	38·27	37·76	37·26	36·79	36·33	35·89	35·47	35·06	34·66
17	40·20	39·64	39·10	38·58	38·08	37·60	37·14	36·70	36·27	35·86
18	41·58	40·99	40·42	39·88	39·36	38·86	38·37	37·91	37·46	37·03
19	42·95	42·33	41·74	41·17	40·62	40·10	39·60	39·11	38·64	38·19
20	44·32	43·67	43·04	42·45	41·88	41·33	40·81	40·30	39·81	39·34
21	*45·68*	45·00	44·35	43·73	43·13	42·56	42·01	41·48	40·98	40·49
22	*47·05*	*46·33*	*45·65*	45·00	44·38	43·78	43·21	42·66	42·13	41·62
23	*48·42*	*47·67*	*46·96*	*46·27*	*45·62*	45·00	44·40	43·83	43·28	42·75
24	*49·80*	*49·01*	*48·26*	*47·55*	*46·87*	*46·22*	*45·60*	45·00	44·43	43·88
25	*51·19*	*50·36*	*49·58*	*48·83*	*48·12*	*47·44*	*46·79*	*46·17*	*45·57*	45·00

Table 6. *Table of values for the Freeman–Tukey square-root transformation**

The observed count is x and the transformed values are $\sqrt{x}+\sqrt{(x+1)}$. The squares of the transformed values are also given for use in analysis of variance computations.

x	$\sqrt{x}+\sqrt{(x+1)}$	$(\sqrt{x}+\sqrt{(x+1)})^2$	x	$\sqrt{x}+\sqrt{(x+1)}$	$(\sqrt{x}+\sqrt{(x+1)})^2$
0	1·00	1·0000	26	10·30	106·0900
1	2·41	5·8081	27	10·49	110·0401
2	3·15	9·9225	28	10·68	114·0624
3	3·73	13·9129	29	10·86	117·9396
4	4·24	17·9776	30	11·04	121·8816
5	4·69	21·9961	31	11·22	125·8884
6	5·10	26·0100	32	11·40	129·9600
7	5·47	29·9209	33	11·58	134·0964
8	5·83	33·9889	34	11·75	138·0625
9	6·16	37·9456	35	11·92	142·0864
10	6·48	41·9904	36	12·08	145·9264
11	6·78	45·9684	37	12·25	150·0625
12	7·07	49·9849	38	12·41	154·0081
13	7·35	54·0225	39	12·57	158·0049
14	7·61	57·9121	40	12·73	162·0529
15	7·87	61·9369	41	12·88	165·8944
16	8·12	65·9344	42	13·04	170·0416
17	8·37	70·0569	43	13·19	173·9761
18	8·60	73·9600	44	13·34	177·9556
19	8·83	77·9689	45	13·49	181·9801
20	9·05	81·9025	46	13·64	186·0496
21	9·27	85·9329	47	13·78	189·8884
22	9·49	90·0601	48	13·93	194·0449
23	9·69	93·8961	49	14·07	197·9649
24	9·90	98·0100	50	14·21	201·9241
25	10·10	102·0100			

* This table was originally published in Gardner Lindzey (editor), *Handbook of Social Psychology*, vol. 1, Addison-Wesley, 1954, p. 327, in a chapter by Frederick Mosteller and Robert R. Rush entitled 'Selected Quantitative Techniques' and is reproduced here with the permission of the Addison-Wesley Publishing Company, Inc.

REFERENCES (*cont.*)

GUPTA, S. S. & SOBEL, M. (1958). On selecting a subset which contains all populations better than a standard. *Ann. Math. Statist.* **29**, 235–44.

MOLINA, E. C. (1942). *Poisson's Exponential Binomial Limit.* Princeton, New Jersey: D. van Nostrand Company, Inc.

MOSTELLER, F. & BUSH, R. R. (1954). Selected quantitative techniques, from *Handbook of Social Pyschology* **1**, 289–334, edited by G. Lindzey. Reading, Mass: Addison-Wesley Publishing Co., Inc.

NATIONAL BUREAU OF STANDARDS (1945). *Table of Arc Sin x.* New York: Columbia University Press.

NATIONAL BUREAU OF STANDARDS (1949). *Tables of the Binomial Probability Distribution.* Applied Mathematics Series **6**, Washington, D.C.: U.S. Government Printing Office.

STEVENS, W. L. (1953). Tables of the angular transformation. *Biometrika*, **40**, 70–3.

U.S. DEPARTMENT OF COMMERCE (1942). *Natural Sines and Cosines to Eight Decimal Places.* Washington, D.C.: U.S. Government Printing Office.

U.S. DEPARTMENT OF COMMERCE (1952). *Tables of the Cumulative Binomial Probabilities.* ORDP 20-1.

Biometrika (1961), **48**, 3 *and* 4, *p.* 441
Printed in Great Britain

Miscellanea

A non-null ranking model for a sequence of m alternatives*

BY C. F. CROUSE
The South African Iron and Steel Industrial Corporation, Pretoria

1. INTRODUCTION

Non-null ranking models, based on the method of paired comparisons, were put forward by Mallows (1957). One of these, called the ϕ-model, was extended by Barton, David & Mallows (1958), to serve as an interpretation and alternative to randomness in a sequence of two alternatives.

In this paper we consider a further extension of this ϕ-model to the case of a sequence of m alternatives.

2. THE ϕ-MODEL

We are given an ordered sequence of $N = \sum_{i=1}^{m} n_i$ elements, $x_{i,j}$ $(j = 1, \ldots, n_i;\ i = 1, \ldots, m)$, and for convenience we will refer to $x_{i,j}$ $(j = 1, \ldots, n_i)$, as a sample from the ith population. (It is assumed that ties do not occur in the ordered sequence.) We are concerned with differences between the populations, and hence distinguish $N!/(n_1! \ldots n_m!)$ different sequences. It is now assumed that this ordered sequence is the outcome of a system of paired comparisons, performed as follows. Each element of every sample is compared independently with every element of every other sample. If following these comparisons, the resulting sequence contains any inconsistencies of the form $x_{i,j} < x_{k,1} < x_{r,s} < x_{i,j}$, it is discarded and a further set of comparisons is performed. This procedure may be supposed to be repeated until a consistent sequence is obtained. As emphasized by Mallows (1957) no experiment would be performed in this fashion, this only being a model for situations where the observed data form a consistent sequence.

It is further assumed that there is no distinction between elements of the same sample so that the paired comparisons are made according to the rule

$$P\{x_{i,a} > x_{j,b}\} = p_{i,j}.$$

Define

$$\phi_{i,j} = p_{i,j}/(1 - p_{i,j}).$$

The null hypothesis, that the $N!/(n_1! \ldots n_m!)$ different sequences are all equally likely, corresponds to

$$\phi_{i,j} = 1, \quad \text{or} \quad p_{i,j} = \tfrac{1}{2} \quad (1 \leqslant i < j \leqslant m), \tag{1}$$

while the ϕ-alternative is given by

$$\phi_{i,j} \neq 1 \quad \text{for at least one } i, j.$$

3. THE DISTRIBUTION OF RANKS UNDER THE ϕ-ALTERNATIVE

Let $U_{i,j}$ be the number of times a $x_{j,r}$ $(r = 1, \ldots, n_j)$, precedes a $x_{i,s}$ $(s = 1, \ldots, n_i)$, in the ranked sequence. Under the ϕ-alternative the probability of obtaining a given consistent sequence, u say, is

$$P(u|\phi) \propto \prod_{i<j} p_{i,j}^{U_{i,j}} (1 - p_{i,j})^{n_i n_j - U_{i,j}}$$

$$= \prod_{i<j} \phi_{i,j}^{U_{i,j}} (1 - p_{i,j})^{n_i n_j}.$$

It follows that under the ϕ-alternative the set $U_{i,j}$, $1 \leqslant i < j \leqslant m$, is sufficient for the set of parameters $\phi_{i,j}$, $1 \leqslant i < j \leqslant m$, and we thus restrict attention to the distribution of the $U_{i,j}$. If $P(U|1)$ is

* Part of Ph.D. thesis submitted and approved, at the University of London.

the probability of obtaining a given set U, say $U_{i,j}$, $1 \leqslant i < j \leqslant m$, under the null hypothesis (1), and $P(U|\phi)$ the corresponding probability under the ϕ-alternative, then it can be shown that

$$P(U|\phi) \propto P(U|1) \prod_{i<j} \phi_{i,j}^{U_{i,j}} (1-p_{i,j})^{n_i n_j},$$

or

$$P(U|\phi) = \frac{P(U|1) \prod_{i<j} \phi_{i,j}^{U_{i,j}}}{\sum_U P(U|1) \prod_{i<j} \phi_{i,j}^{U_{i,j}}}. \tag{2}$$

From (2) the moment generating function of the set U is

$$M(t|\phi) = \frac{\sum_U (\exp \sum_{i<j} t_{i,j} U_{i,j}) P(U|1) \prod_{i<j} \phi_{i,j}^{U_{i,j}}}{\sum_U P(U|1) \prod_{i<j} \phi_{i,j}^{U_{i,j}}}$$

$$= \frac{M(t+\delta|1)}{M(\delta|1)},$$

where

$$t = \{t_{1,2}, t_{1,3}, \ldots, t_{m-1,m}\},$$

$$\delta = \{\delta_{1,2}, \delta_{1,3}, \ldots, \delta_{m-1,m}\}$$

and

$$\delta_{i,j} = \log_e \phi_{i,j}.$$

Thus if $K(t|1)$, and $K(t|\phi)$ are the cumulant generating functions under the respective hypotheses, then

$$K(t+\delta|\phi) = K(t+\delta|1) - K(\delta|1). \tag{3}$$

Let $\kappa^{(\phi)}_{\Pi r_{i,j}}$ and $\kappa^{(1)}_{\Pi r_{i,j}}$ be the respective coefficients of $\prod_{i<j} \frac{t_{i,j}^{r_{i,j}}}{r_{i,j}!}$ in $K(t|\phi)$ and $K(t|1)$. Expanding both sides of (3) and equating the coefficients of $\prod_{i<j} \frac{t_{i,j}^{r_{i,j}}}{r_{i,j}!}$ we obtain

$$\kappa^{(\phi)}_{\Pi r_{i,j}} = \sum_{s_{1,2} \geqslant r_{1,2}} \cdots \sum_{s_{m-1,m} \geqslant r_{m-1,m}} \kappa^{(1)}_{\Pi s_{i,j}} \cdot \frac{(\delta_{1,2})^{s_{1,2}-r_{1,2}}}{(s_{1,2}-r_{1,2})!} \cdots \frac{(\delta_{m-1,m})^{s_{m-1,m}-r_{m-1,m}}}{(s_{m-1,m}-r_{m-1,m})!}.$$

Crouse (1960) has shown that under the null hypothesis the joint distribution of $U_{i,j}/\sqrt{\operatorname{var} U_{i,j}}$, $1 \leqslant i < j \leqslant m$, is asymptotically normal, and furthermore if there exists $\epsilon > 0$ such that

$$1-\epsilon > \lim_{N \to \infty} n_i/(n_i+n_j) > \epsilon$$

for all i, j, then under the null hypothesis the moments of $U_{i,j}/\sqrt{\operatorname{var} U_{i,j}}$ $(1 \leqslant i < j \leqslant m)$, tend to the moments of their limiting distribution with $O(N^{-1})$, and hence as the variance of $U_{i,j}$ under the null hypothesis is

$$(\tfrac{1}{12}) n_i n_j (n_i+n_j+1) = O(N^3),$$

$$\kappa^{(1)}_{\Pi s_{i,j}} = O(N^{\frac{1}{2}\sum_{i<j} s_{i,j} - 1}) \quad \text{for} \quad \sum_{i<j} s_{i,j} \geqslant 3.$$

Thus under the ϕ-alternative

$$\log_e \phi_{i,j} = \delta_{i,j} = c_{i,j}/N^{\frac{1}{2}}, \tag{4}$$

$$\prod_{i<j} [(\tfrac{1}{12}) n_i n_j (n_i+n_j+1)]^{-\frac{1}{2} r_{i,j}} \cdot \kappa_{\Pi r_{i,j}} = O(N^{-1}) \sum_{s_{1,2} \geqslant r_{1,2}} \cdots \sum_{s_{m-1,m} \geqslant r_{m-1,m}} \prod_{i<j} \frac{(c_{i,j})^{s_{i,j}-r_{i,j}}}{(s_{i,j}-r_{i,j})!}$$

for $\sum_{i<j} r_{i,j} \geqslant 3$, and as $\sum_k a^k/k!$ converges for constant a, the above tends to zero when N tends to infinity.
For the lower order moments it is easily found that under (4)

$$\mathscr{E}(U_{i,j}) = \tfrac{1}{2} n_i n_j + N^{-\frac{1}{2}} \sum_{h<k} c_{h,k} \operatorname{cov}_1 (U_{i,j}, U_{h,k}) + O(N^{\frac{1}{2}}), \tag{5}$$

$$\operatorname{cov}(U_{i,j}, U_{h,k}) = \operatorname{cov}_1 (U_{i,j}, U_{h,k}) + O(N^2),$$

where $\mathrm{cov}_1(U_{i,j}, U_{h,k})$, the covariance under the null hypothesis, is equal to

$$(\tfrac{1}{12})n_in_j(n_i+n_j+1) \quad \text{when} \quad i=h, j=k,$$

$$(\tfrac{1}{12})n_in_jn_k \quad \text{when} \quad i=h\neq j\neq k,$$

$$0 \quad \text{when} \quad i\neq j\neq h\neq k.$$

Hence:

THEOREM 1. If as N tends to infinity, there exists $\epsilon > 0$ such that $1-\epsilon > \lim n_i/(n_i+n_j) > \epsilon$, and the limit exists, for all i, j, then under the ϕ-alternative

$$\log_e \phi_{i,j} = c_{i,j}/N^{\frac{1}{2}} \quad (1 \leqslant i < j \leqslant m),$$

the joint distribution of $U_{i,j}/\sqrt{\mathrm{var}_1(U_{i,j})}$, $1 \leqslant i < j \leqslant m$, is asymptotically normal, and apart from a shift in the mean the same as under the null hypothesis.

As shown in unpublished work by Crouse (1960), the rank of the asymptotic distribution of

$$U_{i,j}/\sqrt{\mathrm{var}_1(U_{i,j})} \quad (1 \leqslant i < j \leqslant m),$$

under the null hypothesis, and consequently from Theorem 1 also under (4), is $(m-1)$; the $\binom{m-1}{2}$ linearly independent components for which the asymptotic distribution is degenerate being of the form

$$N^{-\frac{1}{2}}T_{h,i,j} = N^{-\frac{1}{2}}[n_j(U_{h,i}-\tfrac{1}{2}n_hn_i)+n_h(U_{i,j}-\tfrac{1}{2}n_in_j)+n_i(U_{j,h}-\tfrac{1}{2}n_jn_h)].$$

Under the null hypothesis $$N^{-\frac{1}{2}}\mathscr{E}(T_{h,i,j}) = 0,$$

while from (5) it readily follows that under (4)

$$N^{-\frac{1}{2}}\mathscr{E}(T_{h,i,j}) = O(N^{-1}),$$

and the expectation of the degenerate component is thus asymptotically insensitive against (4). Furthermore, under the null hypothesis

$$\mathrm{var}\, T_{h,i,j} = (\tfrac{1}{12})n_hn_in_j(n_h+n_i+n_j),$$

thus even the expectation of the degenerate component standardized under the null hypothesis is asymptotically insensitive against (4).

Finally, it can easily be shown that asymptotically the set $U_{2,j}/\sqrt{\mathrm{var}_1(U_{1,j})}$ $(j = 2, \ldots, m)$, or any similar set, determines the distribution of the set $U_{i,j}/\sqrt{\mathrm{var}_1(U_{i,j})}$, $(1 \leqslant i < j \leqslant m)$, both under the null hypothesis and the alternative (4).

Hence in the limit the problem of testing the null hypothesis against the alternative (4) reduces to: Given a non-singular $(m-1)$-dimensional normal distribution with known variance-covariance matrix, on the basis of a single observed set of values, it is required to test the null hypothesis that all mean values are equal to zero (that is, considering the variables in the form $(U_{i,j}-\frac{1}{2}n_in_j)/\sqrt{\mathrm{var}_1(U_{i,j})}$ against the alternative that at least one mean value is different from zero. This is a special case of the linear hypothesis model (cf. Lehmann, 1959, p. 304) and the corresponding likelihood ratio test statistic, for large samples, is

$$\overline{U}'B^{-1}\overline{U},$$

where $\overline{U}$ is the column vector $(U_{1,2}-\frac{1}{2}n_1n_2, \ldots, U_{1,m}-\frac{1}{2}n_1n_m)$, and B the variance-covariance matrix of this set. Under the hypotheses (1) and (4) the latter can be shown to be asymptotically equivalent to

$$U_m = (12/(N+\tfrac{1}{2}m))\sum_{i<j}(U_{i,j}-\tfrac{1}{2}n_in_j)^2/(n_in_j).$$

Under (1) U_m is asymptotically distributed as a χ^2 with $(m-1)$ degrees of freedom, while under (4) as a non-central χ^2 with $(m-1)$ degrees of freedom and non-centrality

$$\lim_{N\to\infty}(12/(N+\tfrac{1}{2}m))\sum_{i<j}[\mathscr{E}(U_{i,j})-\tfrac{1}{2}n_in_j]^2/(n_in_j).$$

4. CONCLUDING REMARKS

(*a*) For two samples U_m reduces to Wilcoxon's test.

(*b*) Barton *et al.* (1958) have mentioned various situations in which the two-sample test can be applied, and similarly one can give examples for the m-sample test; for instance:

From each of m different nationalities a sample of n_i ($i = 1, \ldots, m$) men is taken; the $N = \sum_{i=1}^{m} n_i$ men all being of the same age. A judge is now requested to arrange the N men in order of increasing age. If the judge is unbiased he should produce a random sequence, a significant value of U_m will thus reflect that men of some nationalities tend to look older than other nationalities.

(*c*) Kruskal (1952) put forward a m-sample extension of Wilcoxon's test, for testing, among other things, the equality of means of m independent samples, and likewise, as will be taken up in a further publication, the test based on U_m can be used for this problem.

The author wishes to express his thanks to Dr F. N. David, Dr C. L. Mallows, and Dr D. E. Barton, for many helpful suggestions.

References

Barton, D. E., David, F. N. & Mallows, C. L. (1958). Non-randomness in a sequence of two alternatives. *Biometrika*, **45**, 166–77.

Crouse, C. F. (1960). Unpublished Ph.D. thesis, University of London.

Kruskal, W. H. (1952). A non-parametric test for the several sample problem. *Ann. Math. Statist.* **23**, 525–40.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* New York: John Wiley and Sons.

Mallows, C. L. (1957). Non-null ranking models. I. *Biometrika*, **44**, 114–30.

Critical values of the coefficient of rank correlation for testing the hypothesis of independence

By Gerald J. Glasser and Robert F. Winter

New York University

A common statistical problem involves the selection of a simple random sample of n bivariate observations in order to test the hypothesis that the two variables are statistically independent. As is well-known, one non-parameteric test of this hypothesis may be based on Spearman's coefficient of rank correlation, r_s, or on an equivalent statistic, D, the sum of the squared differences of the ranks. The relationship between r_s and D is simply

$$r_s = 1 - \frac{6D}{n(n^2-1)}, \quad D = n(n^2-1)(1-r_s). \tag{1}$$

This paper presents approximate critical values of D and r_s for one-tailed α-risks of 0·001, 0·005, 0·010, 0·025, 0·050 and 0·100 with $n = 11\,(1)\,30$ for use in testing the hypothesis of independence. The values are calculated from a Gram–Charlier Type A series approximation to the distribution function of r_s given by David, Kendall & Stuart (1951). The results apply, strictly speaking, only in the case where no tied ranks occur, or where the procedure of breaking ties at random is adopted.

Under the hypothesis of independence the exact probabilities of values of D are based on their relative frequencies in the $n!$ permutations of one ranking against the other. These have been calculated for samples of up to size 10 by Olds (1938), Kendall, Kendall & Babington Smith (1939), and David *et al.* (1951). However, the computations required are extensive and one could not hope to have them carried much further.

For samples with $n > 10$, or at least when $n > 25$ or 30, it is not uncommon in the literature to find a testing procedure based on the asymptotic standardized normality of

$$x_0 = r_s\sqrt{(n-1)}. \tag{2}$$

Olds (1949) gives tables for this purpose. Because D may assume only even values, a correction for continuity may also be introduced by increasing (decreasing) D by unity for values below (above) its mean.

An alternative approximation, suggested by Pitman (1937), consists of approximating the sampling distribution of r_s by a β-distribution, and may be employed by referring to tables of the t-distribution. This leads to the comparison of

$$t_0 = r_s\sqrt{\frac{n-2}{1-r_s^2}} \tag{3}$$

with t^*, a critical value based on the t-distribution with $n-2$ degrees of freedom. This procedure gives

the test exactly the same form as that used in normal theory for the ordinary product moment coefficient of correlation. Tables of critical values based on this approximation has been constructed by Litchfield & Wilcoxon (1955) and by Teegarden (1960). A further correction for continuity in computing r_s for t_0 involves dividing D by $\frac{1}{6}n(n^2-1)+1$ to match the range of the β-distribution.

More elaborate and probably more exact methods of approximating the sampling distribution of r_s, or D, have been given by David *et al.* (1951). The authors give the moments and the cumulants of r_s up to those of the eighth order and use these results in developing the leading terms of an expansion of a distribution function based on Edgeworth's form of the Gram–Charlier Type A series. These results appear to enable one to calculate, with somewhat greater accuracy than either the normal or β-approximations, probabilities of exceeding given values of D or r_s under the hypothesis of independence.

The superiority of the Type A series approximation is pointed up by the results given in Table 1. The table shows several values of $P\{D < D^*\}$ when $n = 10$ and estimates thereof, based on the three approximations mentioned, for values of D^* at probability levels of about 0·001, 0·005, 0·010, 0·025, 0·050, and 0·100. The calculations of the standardized values involved here are straightforward (corrections for the discontinuity of D were employed) and reference is made to the appropriate tables to determine corresponding probabilities. As might be expected, the normal approximation shows up badly at all but the 0·05 level, and the β appears superior. The estimates based on the Type A series are consistently better than either of the other two, although at some levels the β-approximation is almost as good.

Table 1. *Exact and estimated values of* $P\{D < D^*\}$ *for* $n = 10$

	Probabilities				Errors		
D^*	Normal	Beta	Type A	Exact	Normal	Beta	Type A
22	0·00442	0·00059	0·00098	0·00080	+0·00362	−0·00021	+0·00018
24	·00491	·00082	·00134	·00109	+ ·00382	− ·00027	+ ·00025
36	·00904	·00371	·00494	·00439	+ ·00465	− ·00068	+ ·00055
38	·00999	·00456	·00587	·00527	+ ·00472	− ·00071	+ ·00060
44	·01328	·00776	·00926	·00870	+ ·00458	− ·00094	+ ·00056
46	·01456	·00912	·01095	·01012	+ ·00444	− ·00100	+ ·00051
60	·02699	·02346	·02479	·02449	+ ·00250	− ·00103	+ ·00019
62	·02931	·02639	·02744	·02722	+ ·00209	− ·00083	+ ·00008
74	·04716	·04829	·04788	·04814	− ·00098	+ ·00015	− ·00020
76	·05092	·05274	·05218	·05244	− ·00152	+ ·00030	− ·00017
92	·08915	·09838	·09510	·09526	− ·00647	+ ·00276	− ·00024
94	·09527	·10527	·10212	·10222	− ·00695	+ ·00305	− ·00010

The Type A series approximation, despite its apparent excellence, is unfortunately complex and perhaps too burdensome to use in many applications. This paper presents results designed to alleviate this difficulty. The approximation has been used for $n = 11(1)30$ to calculate probabilities for values of D, in the neighbourhood of widely used probability levels. The computations followed along the lines suggested in the paper by David and others, except that Sheppard's corrections for the moments were not introduced (the effect would have been negligible). A correction for discontinuity, however, was employed by deducting unity from each value of D. Several checks for consistency were made.

In computing probabilities, standardized deviates were computed to three decimal places (e.g. 1·649) and probabilities for these deviates to five places (e.g. 0·00486). A more extended calculation might set a few of the critical values given below 2 units higher, or lower, but in no case more. Copies of tables showing the approximated probabilities are available from the authors on request.

From the results, critical values of D have been selected such that, according to the approximation, each is the largest value of D^* satisfying

$$P\{D < D^*\} \leqslant \alpha. \tag{4}$$

Table 2 summarizes these values of D^* for $n > 10$ together with exact critical values for $n \leqslant 10$. Upper-tail critical values of D are obtained from the fact that the distribution is symmetrical on the range $(0, \frac{1}{6}n(n^2-1))$.

A comparison of the values of D^* shown in Table 2 with similar values based on the normal and the β-distributions indicates systematic and, sometimes, substantial differences.

In borderline cases the choice of a table of critical values may determine whether or not the hypothesis of independence is accepted. For example, in his basic statistics textbook, McCarthy (1957) provides an interesting illustration of rank correlation in testing the hypothesis that the success of mediation and the amount of hostility in early labour-management negotiations are independent against the one-tailed alternative that the variables are inversely related. In the illustration $n = 12$ and the results show $D = 454$. Application of the Olds (1949) table based on the normal approximation leads the author to suggest accepting the hypothesis of independence, for $\alpha = 0{\cdot}025$. Table 2, however, shows that the lower 2·5 % level for D is 120, so that the upper 2·5 % level is at $572 - 120 = 452$. The hypothesis would therefore be rejected at this level.

Table 2. *Approximate lower-tail critical values, D^*, where $P\{D < D^*\} \leqslant \alpha$, $n = 4(1)30$*

	Significance level, α					
n	0·001	0·005	0·010	0·025	0·050	0·100
4	—	—	—	—	2	2
5	—	—	2	2	4	6
6	—	2	4	6	8	14
7	2	6	8	14	18	26
8	6	12	16	26	34	44
9	12	22	28	38	50	64
10	22	36	44	60	74	92
11	34	54	66	86	104	128
12	52	78	94	120	144	172
13	76	110	130	162	190	226
14	106	148	172	212	246	290
15	142	194	224	270	312	364
16	186	250	284	340	390	450
17	238	314	356	420	480	550
18	300	390	438	512	582	664
19	372	476	532	618	696	790
20	454	574	638	738	826	934
21	546	686	758	870	972	1092
22	652	810	892	1020	1134	1270
23	772	950	1042	1184	1312	1464
24	904	1104	1208	1366	1510	1678
25	1050	1274	1390	1566	1726	1912
26	1212	1462	1590	1786	1960	2168
27	1390	1666	1808	2024	2216	2444
28	1586	1890	2046	2284	2494	2744
29	1800	2134	2306	2564	2796	3068
30	2032	2398	2584	2868	3120	3416

Note. For the corresponding upper-tail critical values, take $\frac{1}{2}n(n^2-1) - D^*$.

Table 3 presents critical values of the coefficient of rank correlation r_s^*, derived simply from the D^* values in Table 2 by using (1). Lower-tail critical values of r_s are, of course, simply $-r_s^*$.

Table 4 expresses several of the critical values from Tables 2 and 3 in standard units. These figures may be taken to provides some ideas as to the usefulness of the normal approximation outside the range of the table. They indicate, for example, that at the 0·025, 0·050, and 0·100 levels a normal-based test would be almost as accurate as the Type A series approximation when $n > 30$. For smaller levels it would be, supposedly, somewhat less accurate, and the t-test might be applied.

Table 3. *Approximate upper-tail critical values, r_s^*, where $P\{r_s > r_s^*\} \leqslant \alpha$, $n = 4(1)30$*

	Significance level, α					
n	0·001	0·005	0·010	0·025	0·050	0·100
4	—	—	—	—	0·8000	0·8000
5	—	—	0·9000	0·9000	·8000	·7000
6	—	0·9429	0·8857	0·8286	0·7714	0·6000
7	0·9643	·8929	·8571	·7450	·6786	·5357
8	·9286	·8571	·8095	·6905	·5952	·4762
9	·9000	·8167	·7667	·6833	·5833	·4667
10	·8667	·7818	·7333	·6364	·5515	·4424
11	0·8455	0·7545	0·7000	0·6091	0·5273	0·4182
12	·8182	·7273	·6713	·5804	·4965	·3986
13	·7912	·6978	·6429	·5549	·4780	·3791
14	·7670	·6747	·6220	·5341	·4593	·3626
15	·7464	·6536	·6000	·5179	·4429	·3500
16	0·7265	0·6324	0·5824	0·5000	0·4265	0·3382
17	·7083	·6152	·5637	·4853	·4118	·3260
18	·6904	·5975	·5480	·4716	·3994	·3148
19	·6737	·5825	·5333	·4579	·3895	·3070
20	·6586	·5684	·5203	·4451	·3789	·2977
21	0·6455	0·5545	0·5078	0·4351	0·3688	0·2909
22	·6318	·5426	·4963	·4241	·3597	·2829
23	·6186	·5306	·4852	·4150	·3518	·2767
24	·6070	·5200	·4748	·4061	·3435	·2704
25	·5962	·5100	·4654	·3977	·3362	·2646
26	0·5856	0·5002	0·4564	0·3894	0·3299	0·2588
27	·5757	·4915	·4481	·3822	·3236	·2540
28	·5660	·4828	·4401	·3749	·3175	·2490
29	·5567	·4744	·4320	·3685	·3113	·2443
30	·5479	·4665	·4251	·3620	·3059	·2400

Note. The corresponding lower-tail critical value for r_s is $-r_s^*$.

Table 4. *Values of r_s^* in standard deviation units, $n = 10(5)30$*

	Significance level, α					
n	0·001	0·005	0·010	0·025	0·050	0·100
10	2·600	2·345	2·200	1·909	1·655	1·327
15	2·793	2·445	2·245	1·938	1·657	1·310
20	2·871	2·478	2·268	1·940	1·652	1·298
25	2·921	2·498	2·280	1·948	1·647	1·296
30	2·951	2·512	2·289	1·949	1·647	1·293
∞	3·090	2·576	2·326	1·960	1·645	1·282

References

David, S. T., Kendall, M. G. & Stuart, A. (1951). Some questions of distribution in the theory of rank correlation. *Biometrika*, **38**, 131–40.

Kendall, M. G., Kendall, Sheila F. H. & Babington Smith, B. (1939). The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur on equal number of times. *Biometrika*, **30**, 251–73.

Litchfield, J. T., Jr. & Wilcoxon, F. (1955). The rank correlation method. *Analytical Chemistry*, **27**, 299–300.

McCarthy, P. J. (1957). *Introduction to Statistical Reasoning.* New York: McGraw-Hill, Inc.
Olds, E. G. (1938). Distribution of sums of squares of rank difference for small numbers of individuals. *Ann. Math. Statist.* **9**, 133–48.
Olds, E. G. (1949). The 5% significance levels for sums of squares of rank differences and a correction. *Ann. Math. Statist.* **20**, 117.
Pitman, E. J. G. (1937). Significance tests which may be applied to samples from any populations. II. The correlation coefficient test. *J. R. Statist. Soc.*, Suppl. **4**, 225–32.
Pearson, K. (1934). *Tables of the Incomplete Beta-Function.* Cambridge University Press.
Teegarden, K. L. (1960). Critical values for rank order correlations. *Industrial Quality Control*, **16**, 11, 48–49.
U.S. Department of Commerce (1953). *Tables of Normal Probability Functions.* Washington: Government Printing Office.

The bias of the maximum likelihood estimates of the location and scale parameters given a type II censored normal sample

By J. G. SAW

University College London

1. Introduction

Gupta (1952) discussed the solution of the pair of equations in $\hat{\mu}$ and $\hat{\sigma}$, the maximum likelihood estimates of the normal population parameters of location and scale, and gave the leading term for var $(\hat{\mu})$ and var $(\hat{\sigma})$ when it is supposed that only the smallest r of n observations from the parent population are available. The solution of the equations in $\hat{\mu}$ and $\hat{\sigma}$ is tiresome and some attempts have been made to facilitate computations by using a fairly simple, systematic technique; see, for example, Cohen (1957).

It has been convenient to assume that for samples of size $n > 20$ the bias of the likelihood estimates will be small enough as to warrant no consideration. The purpose of this note is to find the leading term (i.e. the term of order $1/n$) in the bias of $\hat{\mu}$ and of $\hat{\sigma}$, first, to comment on the validity of the assumption, and secondly, so that by suitable adjustment the bias may be reduced to order $1/n^2$ in each case.

2. The likelihood equations

Let $x_1 < x_2 < \ldots < x_r$ be the smallest r observations from a sample of size n from a normal population with mean μ and standard deviation σ. We will use

$$z(u) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}u^2}, \quad P(u) = 1 - Q(u) = \int_{-\infty}^{u} z(t)\,dt. \tag{2·1}$$

Then with this notation,

$$p(x_1, x_2, \ldots, x_r) = \frac{n!}{(n-r)!\,\sigma^r} [1 - P((x_r - \mu)/\sigma)]^{n-r} \prod_{i=1}^{r} z((x_i - \mu)/\sigma) \tag{2·2}$$

so that the likelihood equations are

$$\frac{1}{\hat{\sigma}} \sum_{1}^{r} (x_i - \hat{\mu}) + (n-r) \frac{z((x_r - \hat{\mu})/\hat{\sigma})}{Q((x_r - \hat{\mu})/\hat{\sigma})} = 0, \tag{2·3a}$$

$$\frac{1}{\hat{\sigma}^2} \sum_{1}^{r} (x_i - \hat{\mu})^2 + (n-r) \frac{z((x_r - \hat{\mu})/\hat{\sigma})}{Q((x_r - \hat{\mu})/\hat{\sigma})} \frac{(x_r - \hat{\mu})}{\hat{\sigma}} = r. \tag{2·3b}$$

Write

$$x_i = \mu + \sigma v_i, \tag{2·4}$$

then $v_1 < v_2 < \ldots < v_r$ are the r smallest ranked observations of a set of n observations from the standard normal population. We define

$$\bar{v}_r = \sum_{1}^{r} v_i/r, \tag{2·5}$$

$$S_1 = \sum_{1}^{r} (v_i - \bar{v}_r)^2/r, \tag{2·6}$$

$$S_2 = (v_r - \bar{v}_r), \tag{2·7}$$

$$\theta = \hat{\sigma}/\sigma, \tag{2·8}$$

$$u_r = (x_r - \hat{\mu})/\hat{\sigma}, \tag{2·9}$$

$$a_i \quad \text{by} \quad \mathscr{E}(S_i) = a_i + O(1/n) \quad (i = 1, 2), \tag{2·10}$$

$$X_r \quad \text{by} \quad P(X_r) = p_r = r/(n+1), \qquad z_r \quad \text{by} \quad z_r = z(X_r), \tag{2·11}$$

$$\phi_t = (d/dx)^t \left.\frac{z(x)}{Q(x)}\right|_{x=X_r}, \tag{2·12}$$

$$Y(\theta, S_1, S_2) = \frac{1}{\theta}\left[\frac{S_1}{S_2} + S_2\right] - \frac{\theta}{S_2} - X_r, \tag{2·13}$$

$$G(\theta, S_1, S_2) = \frac{S_1}{\theta S_2} - \frac{\theta}{S_2} + \frac{n-r}{r}\sum_{t=0}^{\infty}\frac{\phi_t}{t!}Y^t(\theta, S_1, S_2). \tag{2·14}$$

Now it may be shown (Saw, 1958) that

$$a_1 = 1 - \frac{z_r}{p_r}\left[X_r + \frac{z_r}{p_r}\right], \quad a_2 = X_r + \frac{z_r}{p_r}, \tag{2·15}$$

so that

$$Y(1, a_1, a_2) = 0, \tag{2·16}$$

$$G(1, a_1, a_2) = -z_r/(n+1)\,p_r q_r. \tag{2·17}$$

If we put

$$u_r = X_r + Y(u_r), \tag{2·18}$$

then after expanding $z(u_r)/Q(u_r)$ about $z(X_r)/Q(X_r)$. we have using (2·3a)

$$\frac{1}{\theta}(\bar{v}_r - v_r) + u_r + \frac{n-r}{r}\sum_{t=0}^{\infty}\frac{\phi_t}{t!}Y^t(u_r) = 0. \tag{2·19}$$

Eliminating the factor $z((x_r - \hat{\mu})/\hat{\sigma})/Q((x_r - \hat{\mu})/\hat{\sigma})$ between equations (2·3a) and (2·3b), we arrive after substitution at

$$u_r = \frac{1}{\theta}\left(\frac{S_1}{S_2} + S_2\right) - \frac{\theta}{S_2}, \tag{2·20}$$

so that

$$\begin{aligned} Y(u_r) = u_r - X_r &= \frac{1}{\theta}\left[\frac{S_1}{S_2} + S_2\right] - \frac{\theta}{S_2} - X_r \\ &= Y(\theta, S_1, S_2). \end{aligned} \tag{2·21}$$

Using (2·20) and (2·21), we rewrite equation (2·19) as

$$\frac{S_1}{\theta S_2} - \frac{\theta}{S_2} + \frac{n-r}{r}\sum_{t=0}^{\infty}\frac{\phi_t}{t!}Y^t(\theta, S_1, S_2) = 0, \tag{2·22}$$

or

$$G(\theta, S_1, S_2) = 0. \tag{2·23}$$

Using an inverse Taylor expansion for θ about $\theta = 1$, we have

$$(\theta - 1) = \sum_{k=1}^{\infty}\left[\frac{\partial^k \theta}{\partial G^k}\frac{(-G)^k}{k!}\right]_{(1, S_1, S_2)}. \tag{2·24}$$

In this last equation, θ is to be considered a function of G defined by $G(\theta, S_1, S_2) = 0$.

Expanding the right-hand side of equation (2·24) for S_1 and S_2 about a_1 and a_2 and taking expectations we have finally

$$\mathscr{E}\left(\frac{\hat{\sigma}}{\sigma} - 1\right) = \sum_{i,j=0}^{\infty}\sum h_{i:j} L_{i:j}, \tag{2·25}$$

where

$$h_{i:j} = \sum_{k=1}^{\infty}\left[\left(\frac{\partial}{\partial S_1}\right)^i \left(\frac{\partial}{\partial S_2}\right)^j \frac{\partial^k \theta}{\partial G^k}\frac{(-G)^k}{i!\,j!\,k!}\right]_{(1, a_1, a_2)}, \tag{2·26}$$

$$L_{i:j} = \mathscr{E}(S_1 - a_1)^i (S_2 - a_2)^j \tag{2·27}$$

and it may be shown that $L_{i:j}$ is of order $[\frac{1}{2}(i+j+1)]$ in $1/n$, where $[w]$ is the greatest integer which does not exceed w.

Now by (2·4) and (2·9)

$$(\hat{\mu} - \mu)/\sigma = v_r - \theta u_r. \tag{2·28}$$

Using (2·20) we have after some arrangement

$$(\hat{\mu} - \mu)/\sigma = v_r - S_2 - \frac{S_1 - 1}{S_2} + \frac{(\theta - 1)^2}{S_2} + \frac{2(\theta - 1)}{S_2}. \tag{2·29}$$

3. Moments of S_1 and S_2

In order to determine the coefficient of $1/(n+1)$ in $\mathscr{E}(\hat{\mu})$ and in $\mathscr{E}(\hat{\sigma})$ it is necessary to know the first two moments and product-moment of S_1 and S_2 to order $1/(n+1)$. The technique involved in evaluating these moments in terms of a power series in $1/(n+1)$, the coefficients of which vary with $p_r = r/(n+1)$ only, has been described by David & Johnson (1954) and by the author (1958). The required results are quoted here. If

$$c_i = \frac{1}{i!}\left(\frac{d}{dP(x)}\right)^i x\Big|_{x=X_r}, \quad c_0 = X_r, \tag{3·1a}$$

$$d_i = \frac{1}{i!}\left(\frac{d}{dP(x)}\right)^i \frac{z(x)}{P(x)}\Big|_{x=X_r}, \quad d_0 = z_r/p_r, \tag{3·1b}$$

then

$$\mathscr{E}(S_1-a_1)(n+1) = -p_r q_r(c_0d_2+c_1d_1+c_2d_0+d_1^2+2d_0d_2)+\frac{1}{p_r}[3(c_0+d_0)^2-2(1+c_0d_0+c_0^2)]+O(n)^{-1}, \tag{3·2}$$

$$\mathscr{E}(S_1-a_1)^2(n+1) = p_r q_r(c_0d_1+c_1d_0+2d_0d_1)^2+\frac{1}{p_r}[-2+10d_0^2+11c_0d_0+2c_0^2-10d_0^4-22d_0^3c_0$$
$$-15c_0^2d_0^2-3d_0c_0^3-6(c_0+d_0)^2(1-c_0d_0-d_0^2)+4(1+c_0d_0+c_0^2)(1-c_0d_0-d_0^2)]+O(n)^{-1}, \tag{3·3}$$

$$\mathscr{E}(S_2-a_2)(n+1) = p_r q_r(c_2+d_2)-\frac{1}{p_r}[c_0+d_0]+O(n)^{-1}, \tag{3·4}$$

$$\mathscr{E}(S_2-a_2)^2(n+1) = p_r q_r(c_1+d_1)^2+\frac{1}{p_r}[1-2c_0^2-5c_0d_0-3d_0^2+2(c_0+d_0)^2]+O(n)^{-1}, \tag{3·5}$$

$$\mathscr{E}(S_1-a_1)(S_2-a_2)(n+1) = -p_r q_r(c_1+d_1)(c_0d_1+c_1d_0+2d_0d_1)+\frac{1}{p_r}[c_0^2d_0+2d_0+3c_0+c_0^3$$
$$-5(c_0+d_0)(1-c_0d_0-d_0^2)-3(c_0+d_0)^3+2(c_0+d_0)(1+c_0d_0+c_0^2)]+O(n)^{-1}. \tag{3·6}$$

In addition we have that $\mathscr{E}(S_1-a_1)^i(S_2-a_2)^j$ is of order $[\frac{1}{2}(i+j+1)]$ in $1/(n+1)$.

4. Bias of $\hat{\sigma}$

$G(1, a_1, a_2)$ is itself of order $1/(n+1)$. Using this, we evaluate the coefficients $h_{i:j}$ as follows (each term in the expressions is to be evaluated at the point $(1, a_1, a_2)$)

$$h_{0:0} = -G\frac{\partial\theta}{\partial G}, \quad h_{1:0} = -\frac{\partial G}{\partial S_1}\frac{\partial\theta}{\partial G}+O(n+1)^{-1}, \quad h_{0:1} = -\frac{\partial G}{\partial S_2}\frac{\partial\theta}{\partial G}+O(n+1)^{-1},$$

$$h_{2:0} = \frac{1}{2}\left[-\frac{\partial^2 G}{\partial S_1^2}\frac{\partial\theta}{\partial G}-2\frac{\partial G}{\partial S_1}\frac{\partial^2\theta}{\partial G\,\partial S_1}+\left(\frac{\partial G}{\partial S_1}\right)^2\frac{\partial^2\theta}{\partial G^2}\right]+O(n+1)^{-1},$$

$$h_{0:2} = \frac{1}{2}\left[-\frac{\partial^2 G}{\partial S_2^2}\frac{\partial\theta}{\partial G}-2\frac{\partial G}{\partial S_2}\frac{\partial^2\theta}{\partial G\,\partial S_2}+\left(\frac{\partial G}{\partial S_2}\right)^2\frac{\partial^2\theta}{\partial G^2}\right]+O(n+1)^{-1},$$

$$h_{1:1} = \left[-\frac{\partial^2 G}{\partial S_1\partial S_2}\frac{\partial\theta}{\partial G}-\frac{\partial G}{\partial S_1}\frac{\partial^2\theta}{\partial G\,\partial S_2}-\frac{\partial G}{\partial S_2}\frac{\partial^2\theta}{\partial G\,\partial S_1}+\frac{\partial G}{\partial S_1}\frac{\partial G}{\partial S_2}\frac{\partial^2\theta}{\partial G^2}\right]+O(n+1)^{-1}.$$

Evaluating these leading terms in $h_{i:j}$ and $L_{i:j}$ we arrive, using equation (2·25), at Table 1 which gives the value of $B(\hat{\sigma}: p_r)$ in the expression

$$\mathscr{E}(\hat{\sigma}) = \sigma+\frac{\sigma}{(n+1)}B(\hat{\sigma}: p_r)+O(n+1)^{-2} \tag{4·1}$$

for $p_r = 0{\cdot}20\,(0{\cdot}05)\,0{\cdot}80$.

Table 1. *Values of* $B(\hat{\sigma}: p_r)$, *in* $\mathscr{E}(\hat{\sigma}) = \sigma+\dfrac{\sigma}{n+1}B(\hat{\sigma}: p_r)+O(n+1)^{-2}$

p_r	$B(\hat{\sigma}: p_r)$	p_r	$B(\hat{\sigma}: p_r)$
0·20	−4·777	0·55	−1·582
·25	−3·751	·60	−1·431
·30	−3·087	·65	−1·303
·35	−2·614	·70	−1·193
·40	−2·258	·75	−1·098
·45	−1·983	·80	−1·013
·50	−1·762	1·00	−0·750

5. Bias of $\hat{\mu}$

From (2·29), since

$$\mathscr{E}(v_r) = X_r + \frac{X_r p_r q_r}{(n+1)z_r^2} + O(n+1)^{-2}, \tag{5·1}$$

$$\mathscr{E}\frac{S_1 - 1}{S_2} = \frac{a_1 - 1}{a_2} - \frac{a_1 - 1}{a_2^2}(a_2 L_{0:1} - L_{0:2}) + \frac{1}{a_2^2}(a_2 L_{1:0} - L_{1:1}) + O(n+1)^{-2}, \tag{5·2}$$

$$\mathscr{E}(\theta - 1)^2/S_2 = \frac{\text{var}\,(\theta)}{a_2} + O(n+1)^{-2}, \tag{5·3}$$

$$\mathscr{E}(\theta - 1)/S_2 = \frac{B(\hat{\sigma}:p_r)}{a_2} - \frac{1}{a_2^2}(h_{1:0}L_{1:1} + h_{0:1}L_{0:2}) + O(n+1)^{-2}, \tag{5·4}$$

we may construct Table 2 which gives the value of $B(\hat{\mu}:p_r)$ in the expression

$$\mathscr{E}(\hat{\mu}) = \mu + \frac{\sigma}{(n+1)}B(\hat{\mu}:p_r) + O(n+1)^{-2}. \tag{5·5}$$

Values of $\text{var}\,(\theta) = \text{var}\,(\hat{\sigma}/\sigma)$ can be obtained from Gupta's tables (1952) or from

$$\text{var}\,(\theta) = h_{1:0}^2 L_{2:0} + 2h_{1:0}h_{0:1}L_{1:1} + h_{0:1}^2 L_{0:2} + O(n+1)^{-2}. \tag{5·6}$$

Table 2. *Values of* $B(\hat{\mu}:p_r)$, *in* $\mathscr{E}(\hat{\mu}) = \mu + \dfrac{\sigma}{n+1}B(\hat{\mu}:p_r) + O(n+1)^{-2}$

p_r	$B(\hat{\mu}:p_r)$	p_r	$B(\hat{\mu}:p_r)$
0·20	−5·538	0·55	−0·750
·25	−3·792	·60	− ·583
·30	−2·766	·65	− ·450
·35	−2·078	·70	− ·342
·40	−1·588	·75	− ·254
·45	−1·232	·80	− ·181
·50	−0·960	1·00	− ·000

6. Summary

The tables of $B(\hat{\mu}:p_r)$ and $B(\hat{\sigma}:p_r)$ indicate that the bias in small samples may be considerable when $p_r = r/(n+1)$ is small. For example, for the case $n = 19$, $r = 7$; the bias in the estimate for σ may be as high as 13 % and in the case $n = 19$, $r = 4$; the bias may be as high as 24 %. In order to comment more precisely it would be necessary to obtain some higher terms in the power series expansion in $1/(n+1)$ for the bias of $\hat{\mu}$ and of $\hat{\sigma}$; some additional light could perhaps be thrown by a sampling experiment but in any event it would seem advisable to construct the corrected estimates:

$$\mu_c = \hat{\mu} - \frac{\hat{\sigma}}{(n+1)}B(\hat{\mu}:p_r),$$

$$\sigma_c = \hat{\sigma} - \frac{\hat{\sigma}}{(n+1)}B(\hat{\sigma}:p_r)$$

in order to reduce the bias to order $1/(n+1)^2$.

The research reported here was supported in part by the Department of United States Army and was carried out at the University of North Carolina, Department of Bio-Statistics.

References

Cohen, A. C. (1957). On the solution of estimating equations for truncated and censored samples from normal populations. *Biometrika*, **44**, 225–36.

David, F. N. & Johnson, N. L. (1954). Statistical treatment of censored data. *Biometrika*, **41**, 288–40.

Gupta, A. K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika*, **39**, 260–73.

Saw, J. G. (1958). Moments of sample moments of censored samples from a normal population. *Biometrika*, **45**, 211–21.

On the solution of the likelihood equation by iteration processes

BY B. K. KALE

Department of Mathematics and Statistics, University of Poona, India

1. The exact determination of the maximum likelihood estimate (MLE) is often quite difficult. In such cases MLE is approximated by using some iteration process, the usual being the method of scoring for parameters as given by Rao (1952, pp. 165–7). The idea is essentially due to Fisher (1925).

If T_1 is the trial solution of the likelihood equation, T_r the rth iterate and $\hat{\theta}$ the MLE, then the use of an iteration process is justified if the error $|T_r - \hat{\theta}|$ decreases with increased iterations and tends to zero as r tends to infinity. As T_r and $\hat{\theta}$ are both random variables, we have to investigate the probability with which the above properties are realized in an application of an iteration process. The use of the iteration process is justified in large samples at least, if the above probability tends to one as n, the sample size, tends to infinity.

In this paper we consider some iteration processes, and the sufficient conditions under which they have the desirable properties. We afterwards investigate the probability with which the sufficient conditions are satisfied.

In effect, it is shown that the iteration processes usually applied in practice are justifiable, in large samples at least.

2. Let $f(x, \theta)$ be the probability density function which satisfies the regularity conditions given by Cramér (1954), and whose range does not depend upon the parameter. In this paper we will consider the single-parameter case only; the multiparametric case will be considered in a separate paper.

Under the regularity conditions, the following well-known results have already been proved by Cramér (1954) and Huzurbazar (1948).

R. 1. With probability approaching certainty as $n \to \infty$, the likelihood equation admits a consistent solution (Cramér).

R. 2. $(1/n)(d^2 \log L/d\theta^2)_{\theta=\hat{\theta}}$ converges in probability to $-I(\theta_0)$ as $n \to \infty$ (Huzurbazar).

R. 3. The consistent solution of the likelihood equation is unique and $P[(d^2 \log L/d\theta^2)_{\theta=\hat{\theta}} < 0] \to 1$ as $n \to \infty$ (Huzurbazar).

The following lemma will be used in the sequel:

LEMMA. If Y and Z are random variables and a and b are positive constants,

$$P[|Y+Z| \geqslant a+b] \leqslant P[|Y| \geqslant a] + P[|Z| \geqslant b].$$

This result has been given by Frechét (1950, chapter 1).

3. Let $\psi(\theta)$ be a differentiable function of θ, which has no zero in a neighbourhood of $\hat{\theta}$, the root of likelihood equation which we assume to exist.

Define

$$\phi(\theta) = \theta - \psi(\theta) \frac{d \log L}{d\theta}. \tag{1}$$

Consider the iteration process $T_{r+1} = [\phi(\theta)]_{\theta=T_r}$

$$T_{r+1} = T_r - \psi(T_r) \left(\frac{d \log L}{d\theta}\right)_{\theta=T_r}. \tag{2}$$

Let $e_r = |T_r - \hat{\theta}|$ be the error at the rth iteration, then the choice of $\psi(\theta)$ is to be made in such a way that $e_{r+1} < e_r$ and $e_r \to 0$ as $r \to \infty$, so that process (2) converges to $\hat{\theta}$.

Householder (1953, pp. 118–22) has shown that the following conditions are sufficient to ensure that $e_{r+1} < e_r$ and that $e_r \to 0$.

(A) There exists a ρ neighbourhood of $\hat{\theta}$, $N_\rho(\hat{\theta})$, such that if θ' and θ'' ϵ $N_\rho(\hat{\theta})$ we have for some $k \geqslant 0$ $|\phi(\theta') - \phi(\theta'')| \leqslant k$, where $0 \leqslant k < 1$. If further $\phi(\theta)$ is differentiable, $|\phi'(\hat{\theta})| < 1$ ensures the existence of $N_\rho(\hat{\theta})$.

(B) The initial solution $T_1 \epsilon N_\rho(\hat{\theta})$.

We will consider different iteration processes corresponding to the different choices of $\psi(\theta)$, and then investigate with what probability the conditions (A) and (B) are satisfied. If, at least, we can prove that conditions (A) and (B) are satisfied with probability approaching certainty as $n \to \infty$, the use of the iteration process is justified in large samples.

4. Consider first the Newton–Raphson process, where $\psi(\theta) = [d^2 \log L/d\theta^2]^{-1}$. As by R. 3

$$P\left[\left(\frac{d^2 \log L}{d\theta^2}\right)_{\theta=\hat{\theta}} < 0\right] \to 1 \quad \text{as} \quad n \to \infty,$$

and $[d^2 \log L/d\theta^2]^{-1}$ is differentiable for θ, $\psi(\theta)$ has no zero in some neighbourhood of $\hat{\theta}$ with probability approaching unity as $n \to \infty$.

$$\phi(\theta) = \theta - \frac{d \log L}{d\theta} \Big/ \frac{d^2 \log L}{d\theta^2},$$

therefore $\phi'(\hat{\theta}) = 0$, on the assumptions that $(d \log L/d\theta)_{\theta=\hat{\theta}} = 0$. By R. 1 $\hat{\theta}$, exists with probability approaching unity as $n \to \infty$. As $\phi'(\hat{\theta}) = 0$ implies $|\phi'(\hat{\theta})| < 1$, the probability that condition (A) is satisfied, tends to unity as $n \to \infty$.

As for condition (B), we will choose T_1 to be a consistent estimate; then as T_1 and $\hat{\theta}$ are both consistent.

$$P[|T_1 - \theta_0| \geqslant \tfrac{1}{2}\rho] \to 0 \quad \text{as} \quad n \to \infty, \tag{3}$$

and

$$P[|\hat{\theta} - \theta_0| \geqslant \tfrac{1}{2}\rho] \to 0 \quad \text{as} \quad n \to \infty. \tag{4}$$

Now $$T_1 - \hat{\theta} = (T_1 - \theta_0) + (\theta_0 - \hat{\theta}).$$

Hence applying Frechet's lemma

$$P[|T_1 - \hat{\theta}| \geqslant \rho] \leqslant P[|T_1 - \theta_0| \geqslant \tfrac{1}{2}\rho] + P[|-\theta_0|\hat{\theta} \geqslant \tfrac{1}{2}\rho].$$

By (3) and (4) it follows that $P[|T_1 - \hat{\theta}| \geqslant \rho] \to 0 \quad \text{as} \quad n \to \infty.$

Thus

$$P[T_1 \in N_\rho(\hat{\theta})] \to 1 \quad \text{as} \quad n \to \infty, \tag{5}$$

when T_1 is consistent. This shows that condition (B) is also satisfied with probability approaching unity as $n \to \infty$.

In all the other processes also, the initial solution will be chosen to be consistent, and hence by (5) the probability, that condition (B) is satisfied, approaches certainty as $n \to \infty$. This result we will take as proved one in other processes and we will concentrate attention only on condition (A).

5. Now consider the method of scoring for parameters. Here

$$\psi(\theta) = -\frac{1}{nI(\theta)}, \quad \text{where} \quad I(\theta) = E\left(-\frac{d^2 \log f}{d\theta^2}\right) \quad \text{and} \quad 0 < I(\theta) < \infty.$$

Now $[I(\theta)]^{-1}$ has no zero in the neighbourhood of $\hat{\theta}$. We further assume that $I(\theta)$ is differentiable. We have

$$\phi(\theta) = \theta + \frac{d \log L}{d\theta} \Big/ nI(\theta),$$

therefore

$$\phi'(\hat{\theta}) = 1 + \left[\frac{d^2 \log L}{d\theta^2} \Big/ nI(\theta)\right]_{\theta=\hat{\theta}} \quad \text{as} \quad \left(\frac{d \log L}{d\theta}\right)_{\theta=\hat{\theta}} = 0.$$

The condition $|\phi'(\hat{\theta})| < 1$ is equivalent to

$$-2I(\hat{\theta}) < \frac{1}{n}\left(\frac{d^2 \log L}{d\theta^2}\right)_{\theta=\hat{\theta}} < 0. \tag{6}$$

By R. 3,

$$P\left[\frac{1}{n}\left(\frac{d^2 \log L}{d\theta^2}\right)_{\theta=\hat{\theta}} < 0\right] \to 1 \quad \text{as} \quad n \to \infty;$$

hence the right-hand part of the inequality (6) is satisfied with probability approaching unity as $n \to \infty$.

As for the other part, consider $(1/n)(d^2 \log L/d\theta^2)_{\theta=\hat{\theta}} + 2I(\hat{\theta})$. From R. 2 $(1/n)(d^2 \log L/d\theta^2)_{\theta=\hat{\theta}}$ converges in probability to $-I(\theta_0)$ as $n \to \infty$. As $I(\theta)$ is differentiable, $I(\theta)$ is also a continuous function of θ. Further $\hat{\theta}$ converges in probability to θ_0 as $n \to \infty$. Hence $I(\hat{\theta})$ converges in probability to $I(\theta_0)$ as $n \to \infty$. This follows from the result given by Frechet (1950), that if x_n converges in probability to x then if g is a continuous function $g(x_n)$ converges in probability to $g(x)$. Thus $(1/n)(d^2 \log L/d\theta^2)_{\theta=\hat{\theta}} + 2I(\hat{\theta})$ converges in probability to $I(\theta_0)$ which is positive. Therefore

$$\left.\begin{aligned} P\left[\frac{1}{n}\left(\frac{d^2 \log L}{d\theta^2}\right)_{\theta=\hat{\theta}} + 2I(\hat{\theta}) > 0\right] \to 1 \quad \text{as} \quad n \to \infty, \\ P\left[-2I(\hat{\theta}) < \frac{1}{n}\left(\frac{d^2 \log L}{d\theta^2}\right)_{\theta=\hat{\theta}}\right] \to 1 \quad \text{as} \quad n \to \infty. \end{aligned}\right\} \tag{7}$$

This shows that left-hand part of inequality (6) is also satisfied with probability approaching certainty as $n \to \infty$. Thus the condition (A) is satisfied with probability approaching unity as $n \to \infty$.

6. The choice $\psi(\theta) = -k/n$ for a suitable choice of constant k is very convenient as the numerical calculations are reduced. For this choice

$$\phi(\theta) = \theta + \frac{k}{n}\frac{d\log L}{d\theta} \quad \text{and} \quad \phi'(\hat{\theta}) = 1 + \frac{k}{n}\left(\frac{d^2\log L}{d\theta^2}\right)_{\theta=\hat{\theta}}$$

and the condition equivalent to (A) is

$$-2 < \frac{k}{n}\left(\frac{d^2\log L}{d\theta^2}\right)_{\theta=\hat{\theta}} < 0. \tag{8}$$

Choose $k > 0$ then by R. 3

$$P\left[\frac{k}{n}\left(\frac{d^2\log L}{d\theta^2}\right)_{\theta=\hat{\theta}} < 0\right] \to 1 \quad \text{as} \quad n \to \infty.$$

Now $(k/n)(d^2\log L/d\theta^2)_{\theta=\hat{\theta}} + 2$ converges in probability to $-kI(\theta_0) + 2$ as $n \to \infty$. If $I(\theta_0)$ is known we can choose $k > 0$ so that $-kI(\theta_0) + 2 > 0$ so that inequality (8) is satisfied with probability approaching unity as $n \to \infty$. For example, in the case of Cauchy's distribution given by $1/[\pi\{1 + (x-\theta)^2\}]$, $I(\theta_0) = \frac{1}{2}$ so that any k from $(0, 4)$ can be chosen. The best choice is $k = 2$ for which $\phi'(\hat{\theta})$ converges in probability to zero, and it is known that for such a choice of $\phi(\theta)$, the process is rapidly convergent.

Such a choice is possible if $I(\theta)$ is constant, which is very rarely the case. As suggested by Huzurbazar (1955), we may first try the parameter transformation θ to α which will preserve regularity conditions and will make $I(\theta)$ independent of the parameter, and then apply the process with suitable choice of k.

7. Define the iteration process as follows,

$$T_{r+1} = T_r + \frac{a_r}{n}\left(\frac{d\log L}{d\theta}\right)_{\theta=T_r}, \tag{9}$$

where a_r is a sequence of real numbers to be suitably chosen, such that $e_{r+1} < e_r$ and $e_r \to 0$ as $r \to \infty$.

Hildebrand (1956, pp. 443–50) has shown that for suitable choice of a_r the following conditions are sufficient

(A) $\quad -2 < \dfrac{a_r}{n}\left(\dfrac{d^2\log L}{d\theta^2}\right)_{\theta=\hat{\theta}} < 0 \quad$ for all $r \geqslant r_0$,

(B) $\quad T_1 \in N_\rho(\hat{\theta})$.

For fixed r, $(a_r/n)(d^2\log L/d\theta^2)_{\theta=\hat{\theta}}$ converges in probability to $-I(\theta_0)a_r$. Therefore, if we choose $a_r > 0$ and such that $a_r \to 0$ as $r \to \infty$, then as $d^2\log L/d\theta^2$ is almost everywhere finite, there exists r_0 such that

$$-2 < \frac{a_r}{n}\left(\frac{d^2\log L}{d\theta^2}\right)_{\theta=\hat{\theta}} \quad \text{for all } r \geqslant r_0,$$

and as $a_r > 0$ we have

$$P\left[-2 < \frac{a_r}{n}\left(\frac{d^2\log L}{d\theta^2}\right)_{\theta=\hat{\theta}} < 0 \quad \text{for all } r \geqslant r_0\right] \to 1 \quad \text{as} \quad n \to \infty.$$

Thus the probability that conditions (A) is satisfied approaches certainty as $n \to \infty$. Again for T_1 we use a consistent estimate so that the condition (B) is also satisfied with probability approaching certainty as $n \to \infty$.

8. If a distribution admits a sufficient statistic, the likelihood equation has the form

$$\Phi(\theta) = \sum_{i=1}^{n} u(x_i), \quad \text{i.e.} \quad \theta = \Phi^{-1}\left[\sum_{i=1}^{n} u(x_i)\right]. \tag{10}$$

We may then apply any of the iteration processes considered above to solve (10). If the iteration process is applied, the condition (A) is satisfied with certainty even for finite n, but the condition (B) is satisfied with probability approaching certainty as $n \to \infty$, and the results are not free from probability.

Instead we may use inverse interpolation methods as follows. We calculate several values $\Phi(\theta_1)$, $\Phi(\theta_2)$, ..., $\Phi(\theta_m)$, where (θ_1, θ_m) encloses $\hat{\theta}$, and then interpolate by using a suitable formula for the observed value $\sum_{i=1}^{n} u(x_i)$. No considerations of probability will arise in such a procedure. The methods of

inverse interpolation can, however, be only applied conveniently, when the likelihood equation has form (10), in which case it can be easily seen that the distribution admits a sufficient statistic.

For example, consider the estimation of λ where

$$f(x,\lambda) = \frac{1}{\Gamma(\lambda+1)}\, e^{-x} x^{\lambda} \quad (x \geqslant 0,\ \lambda+1 \geqslant 0).$$

Here $$\Phi(\lambda) = n\frac{d}{d\lambda}[\log \Gamma(\lambda+1)] \quad \text{and} \quad u(x) = \log x.$$

Φ^{-1} is very complicated and an explicit solution for $\hat{\lambda}$ is not possible. We may use Pairman's (1919) *Tables of Digamma Function* to tabulate $\Phi(\lambda)$ for various values of λ and then interpolate for $\sum_{i=1}^{n} u(x_i)$.

9. By way of comparison of the iterative processes, only the Newton–Raphson process is of second order, while all the others are of first order. The Newton–Raphson process is accordingly rapidly convergent. However, this process usually involves more cumbersome calculations than any other. The method of scoring for parameters can be applied if $I(\theta)$ is differentiable for θ, which is usually the case. The choice of $\psi(\theta)$ to be a suitable constant is very rarely possible. As for the selection of suitable sequence a_r in the process given by (9), we must have $a_r \to 0$ as $r \to \infty$ and $a_r > 0$. However, in a given problem the choice of a_r is not as arbitrary as it might appear, if the method is to be practically useful. It is to some extent governed by the behaviour of $d \log L/d\theta$, as it should be. We should choose a_r so that the process does not converge too slowly. A simple choice of k/r or k/r^2 is often quite convenient.

To compare the methods we will consider the following example given by Fisher (1954, pp. 299–320), based on Carver's data, for two factors in maize, starchy *v.* sugary and green *v.* white. The problem is to evaluate the maximum likelihood estimate of θ, when the probability of belonging to four classes, respectively, is

$$\tfrac{1}{4}(2+\theta), \quad \tfrac{1}{4}(1-\theta), \quad \tfrac{1}{4}(1-\theta), \quad \tfrac{1}{4}\theta.$$

Table 1

Starchy		Sugary		
Green	White	Green	White	Total
1997	906	904	32	3839

As shown by Fisher, quite a few consistent estimates are available. We use $T_1 = (1/n)(a-b-c+d)$, where a, b, c, d are the observed frequencies of the respective classes and n is the total number of cases. Thus $T_1 = 0{\cdot}057046$. Fisher has shown that the likelihood equation is quadratic and has two roots, one positive and one negative. The positive root is the required value of $\hat{\theta}$, namely 0·035712. As the exact value of $\hat{\theta}$ was available, the comparison of the methods was easier.

10. The following table shows the successive iterates obtained by three methods:

(I) Newton–Raphson process;

(II) method of scoring for parameters;

(III) the choice of an arbitrary sequence a_r.

The choice of $a_r = 1/(5r)$ in (III) was based on the observation that $(1/n)(d \log L/d\theta)_{\theta = T_1} = -0{\cdot}100997$ and that the likelihood equation suggested that $\hat{\theta} < T_1$, but $\hat{\theta} > 0{\cdot}03$.

Four iterations are calculated. The entry 'Adjustment' gives the value of correction to T_4. The 'Error' gives the value of $|T_5 - \hat{\theta}|$. The calculations for (I) were found in an interesting note by Norton (1956). Except T_4, our values differ from those of Norton in the sixth place of decimals. T_2 and T_3 for method (II) were given by Stuart (1958). Our value of T_3 differs from that of Stuart at the fourth place of decimals.

Table 2. *Successive iterations by three different methods listed above*

Iterates	(I)	(II)	(III)
T_1	0·057046	0·057046	0·057046
T_2	·025631	·036986	·036847
T_3	·033004	·035791	·036057
T_4	·035707	·035718	·035894
T_5	·035713	·035713	·035832
Adjustment	·000006	−·000005	−·000062
Error	·000001	·000001	·000120

It has been observed in § 9 that method I is rapidly convergent and is asymptotically better than method II. The above example, however, indicates that method II is closer to the true value than method I, for the two iterations T_2 and T_3, and as good thereafter. This needs some explanation.

An iteration process is of order k if $\lim_{r\to\infty}(e_{r+1})/\{(e_r)^k\} = \alpha$, where α is a constant depending on the root of the equation; thus for large r, for the kth-order iteration process

$$e_{r+1} \sim (e_r)^k \alpha(\theta). \tag{11}$$

When the process is convergent $e_r < 1$ and the higher the order k, the more rapidly will $e_r \to 0$ as $r \to \infty$. For method I we have

$$k = 2, \quad \alpha_1(\theta) = \left| \frac{\frac{1}{2}\left(\frac{d^3 \log L}{d\theta^3}\right)_{\theta=\hat{\theta}}}{\left(\frac{d^2 \log L}{d\theta^2}\right)_{\theta=\hat{\theta}}} \right| \tag{12}$$

and for method II we have

$$k = 1, \quad \alpha_2(\theta) = \left| \frac{1 + \left(\frac{d^2 \log L}{d\theta^2}\right)_{\theta=\hat{\theta}}}{nI(\hat{\theta})} \right|. \tag{13}$$

Method I will be therefore generally better than method II, unless $\alpha_2(\hat{\theta})$ is slightly smaller than $\alpha_1(\hat{\theta})$. If $\alpha_2(\hat{\theta})$ is substantially smaller than $\alpha_1(\hat{\theta})$, Method II may give better results for a few iterations. In the example considered above, $\alpha_1(\hat{\theta}) = 25{\cdot}55$ while $\alpha_2(\hat{\theta}) = 0{\cdot}06$.

Note that as $\alpha_2(\hat{\theta})$ converges in probability to zero as $n \to \infty$, for large samples $\alpha_2(\hat{\theta})$ will be usually small. Moreover, as $\hat{\theta}$ is not known, the choice between the methods may be based on $\alpha_1(T_1)$ and $\alpha_2(T_1)$. If $\alpha_1(T_1)$ is substantially larger than $\alpha_2(T_1)$, there may not be much advantage in choosing method I as its application involves rather laborious calculations. In the example given above $\alpha_1(T_1) = 16{\cdot}36$ and $\alpha_2(T_1) = 0{\cdot}373$, and method II may be preferred.

I sincerely thank Dr V. S. Huzurbazar for his guidance and encouragement in preparing this paper and Dr B. R. Rao for going through an earlier draft. I also thank the referee for drawing my attention to Stuart's paper. I thank the Government of India for awarding me a research training scholarship.

References

Cramér, H. (1954). *Mathematical Methods of Statistics* (6th printing). Princeton University Press.

Fisher, R. A. (1925). Theory of statistical estimation. *Proc. Camb. Phil. Soc.* **22**, 700–25.

Fisher, R. A. (1954). *Statistical Methods for Research Workers* (13th edition). Edinburgh: Oliver and Boyd, Ltd.

Frechét, M. (1950). *Recherches Théoriques Modernes sur la Théorie des probabilités*, Première Livre. Paris: Gauthier-Villars.

Hildebrand, F. B. (1956). *Introduction to Numerical Analysis*. New York: McGraw-Hill Book Co.

Householder, A. S. (1953). *Principles of Numerical Analysis*. New York: McGraw-Hill Book Co.

Huzurbazar, V. S. (1948). The likelihood equation, consistency and the maxima of the likelihood function. *Ann. Eugen., Lond.*, **14**, 185–200.

Huzurbazar, V. S. (1955). On successive approximation to the maximum likelihood estimator. *Poona Univ. J. Sci. Tech.* **12**, 96A–97.

Norton, H. W. (1956). One likelihood adjustment may be inadequate. *Biometrics*, **12**, 79–81.

Pairman, Eleanor (1919). *Tables of the Digamma and Trigamma Functions*. Tracts for Computers, 1. Cambridge University Press.

Rao, C. R. (1952). *Advanced Statistical Methods for Biometric Research*. New York: John Wiley and Sons.

Stuart, A. (1958). Iterative solutions of likelihood equations. *Biometrics*, **14**, 128–30.

Detection of best and outlying normal populations with known variances

By A. ZINGER

University of Montreal, Canada

1. Summary

The following two problems are considered in the case of several normal populations with known variances:

(i) The selection of the 'best' population (the population with the largest mean), given a probability of taking a wrong decision.

(ii) A test for outliers in the case of up to seven normal populations.

The statistic used in both cases is the standardized difference between the two largest sample means. The 'best' population can also be defined as the one with the smallest mean.

2. Introduction

The present paper is an extension of a similar problem considered by Zinger & St-Pierre (1958) in the case of three normal populations.

The non-null distribution of the standardized difference between the two largest sample values is derived in the general case of several normal populations with known variances. The least favourable configurations are given and two tests are presented. One is for detecting the 'best' population and does not depend on the number of populations. The other one is for detecting outliers and some critical values are given for up to seven normal populations. All numerical calculations were done on an LGP-30 computer.

3. Non-null distribution of the standardized difference between the two largest sample means

Consider $n+1$ normal populations with unknown means μ_i and known variances $\sigma_i^2 (i = 0, 1, \ldots, n)$. From the ith population, a sample of n_i values $x_{ij} (i = 0, 1, \ldots, n; j = 1, \ldots, n_i)$, is drawn. Let $\bar{x}_i$ be the sample mean associated with the population having mean μ_i. Let $\bar{x}_{(0)} > \bar{x}_{(1)} > \ldots > \bar{x}_{(n)}$ be the ordered sample means; the event $\bar{x}_i = \bar{x}_j$, $i \neq j$, will of course be neglected. Consider $\mu_{(0)} > \mu_{(1)} > \ldots > \mu_{(n)}$, the ordered unknown means. It is not known which population is associated with $\mu_{(i)}$. Only the following $(n+1)!$ mutually exclusive events are possible: $(\bar{x}_{(0)}, \bar{x}_{(1)}, \ldots, \bar{x}_{(n)})$ comes from the populations with means $(\mu_{(i_0)}, \mu_{(i_1)}, \ldots, \mu_{(i_n)})$, with $i_0 \neq i_1 \neq \ldots \neq i_n = 0, 1, \ldots, n$. It follows that the joint density of $\bar{x}_{(0)}, \bar{x}_{(1)}, \ldots, \bar{x}_{(n)}$ is given by

$$f(\bar{x}_{(0)}, \bar{x}_{(1)}, \ldots, \bar{x}_{(n)}) = \prod_{i=0}^{n} \frac{n_i^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}\sigma_i} \Sigma' \exp\left\{-\frac{1}{2}\sum_{k=0}^{n} \frac{n_{i_k}(\bar{x}_{(k)} - \mu_{(i_k)}^2)}{\sigma_{ik}^2}\right\},$$

where Σ' stands for summation over all permutations of $i_0, i_1, \ldots, i_n$. The proposed tests assume that the values of $\sigma_0, \sigma_1, \ldots, \sigma_n$ are known and that the sample sizes are chosen so that $\sigma_i^2/n_i = \sigma^2$ (say 1) $(i = 0, 1, \ldots, n)$.

To obtain the distribution of

$$u = (\bar{x}_{(0)} - \bar{x}_{(1)}),$$

let

$$y_0 = (\bar{x}_{(0)} - \mu_{(i_0)}) + \ldots + (\bar{x}_{(n)} - \mu_{(i_n)})$$

and

$$y_k = (\bar{x}_{(0)} - \mu_{(i_0)}) - (\bar{x}_{(k)} - \mu_{(i_k)}) \quad (k = 1, \ldots, n).$$

Integrating out y_0, one gets

$$f(y_1, \ldots, y_n) = \frac{1}{(n+1)^{\frac{1}{2}}(2\pi)^{\frac{1}{2}n}} \exp\left\{-\frac{1}{2}\frac{1}{n+1}\left(n\sum_{i=1}^{n} y_i^2 - 2\sum_{1 \leqslant i < j} y_i y_j\right)\right\}.$$

The domain of integration becomes

$$y_k > y_{k-1} + \mu_{(i_k)} - \mu_{(i_{k-1})} \quad (k = 2, \ldots, n);$$

and

$$f(y_1) = \frac{1}{(n+1)^{\frac{1}{2}}(2\pi)^{\frac{1}{2}n}} \int_{y_1 + \mu_{(i_2)} - \mu_{(i_1)}}^{\infty} \ldots \int_{y_{n-1} + \mu_{(i_n)} - \mu_{(i_{n-1})}}^{\infty} f(y_1, \ldots, y_n)\, dy_n \ldots dy_2.$$

Let $u_1 = y_1$ and

$$u_k = y_k \frac{\sqrt{k}}{\sqrt{(k+1)}} - \frac{1}{\sqrt{k}\sqrt{(k+1)}}(y_1 + \ldots + y_{k-1}) \quad (k = 2, \ldots, n).$$

The domain of integration becomes for $k = 3, \ldots, n$

$$u_k > \frac{\sqrt{(k-1)}\, u_{k-1} + \sqrt{k}(\mu_{(i_k)} - \mu_{(i_{k-1})})}{\sqrt{(k+1)}}$$

and

$$u_2 > \frac{u_1 + 2(\mu_{(i_2)} - \mu_{(i_1)})}{\sqrt{6}}.$$

Putting $u = u_1 - (\mu_{(i_1)} - \mu_{(i_0)})$, one gets

$$f(u) = \frac{1}{\sqrt{2}\,(2\pi)^{\frac{1}{2}n}} \Sigma' \exp \tfrac{1}{2}\{-(u + \mu_{(i_1)} - \mu_{(i_0)})^2\} \cdot$$

$$\times \int_{(u - \mu_{(i_0)} - (\mu_{(i_1)} + 2\mu_{(i_2)})/\sqrt{6}}^{\infty} \exp(-\tfrac{1}{2}u_2^2) \cdot$$

$$\times \int_{(\sqrt{2}\,u_2 + \sqrt{3}(\mu_{(i_2)} - \mu_{(i_3)}))/\sqrt{4}}^{\infty} \exp(-\tfrac{1}{2}u_3^2) \ldots$$

$$\ldots \int_{(\sqrt{(n-1)}\,u_{n-1} + \sqrt{n}(\mu_{(i_n)} - \mu_{(i_{n-1})}))/\sqrt{(n+1)}}^{\infty} \exp(-\tfrac{1}{2}u_n^2)\, du_n \ldots du_2.$$

The function $f(u)$, consisting of $(n+1)!$ terms, can be written as the sum of a function $g(u)$, containing the $n!$ terms with $i_0 = 0$, which correspond to a good decision, and of a function $w(u)$, containing the remaining $n \cdot n!$ terms with $i_0 \neq 0$, which correspond to a wrong decision.

4. Detection of the best population

In order to detect the best population the following procedure is proposed:

(i) Draw n_i observations from the ith population, subject to the restrictions

$$n_i/\sigma_i^2 = 1/\sigma^2 \quad (i = 0, 1, \ldots, n).$$

(ii) Let

$$\bar{x}_{(0)} > \bar{x}_{(1)} > \ldots > \bar{x}_{(n)}$$

be the ordered sample means. Compute $u = (\bar{x}_{(0)} - \bar{x}_{(1)})/\sigma$.

(iii) If $u \geqslant k$, decide that $\bar{x}_{(0)}$ comes from the population with means $\mu_{(0)}$;

if $u < k$ do not take the above decision.

This procedure will be referred to as the I-procedure, since it was introduced by Irwin (1925).

The critical value k is to be chosen in such a way that the probability of a wrong decision is at most equal to a number $\alpha (0 < \alpha < 1)$ given in advance. α will be called the level. Define

$$\gamma = (\mu_{(0)} - \mu_{(1)})/\sigma$$

and

$$\delta_i = (\mu_{(i)} - \mu_{(i+1)})/\sigma \quad (i = 1, \ldots, n-1).$$

If

$$W(k, \gamma, \delta_1, \ldots, \delta_{n-1}) = \int_k^{\infty} w(u)\, du,$$

then the condition becomes $\Pr(u \geqslant k$ and $\bar{x}_{(0)}$ comes from $\mu_{(1)}$ or $\ldots$ or $\mu_{(n)}) = W(k, \gamma, \delta_1, \ldots, \delta_n) \leqslant \alpha$.

There remains the choice of the least favourable configuration, i.e. of the least favourable value of the vector $(\gamma, \delta_1, \ldots, \delta_n)$. This least favourable configuration is not unique. By numerical calculations it is possible to show that for a given $n+1$ there exists a value k_0 and a corresponding level α_0 such that:

If $k \leqslant k_0$, the least favourable configuration is the null configuration $(0, 0, \ldots, 0)$; if $k \geqslant k_0$, the least favourable configuration is the pseudo-null configuration $(0, \infty, \ldots, \infty)$.

The values of k_0 and α_0 for $n+1 = 3, \ldots, 7$ are:

Table 1

$n+1$	k_0	α_0
3	0·8568	0·2723
4	·7537	·2970
5	·6906	·3127
6	·6468	·3237
7	·6142	·3320

Since, in practice, the levels will always be smaller than α_0, the value k is solution of $W(k, 0, \infty \ldots, \infty) = \alpha$, i.e.

$$\frac{1}{\sqrt{2\pi}} \int_{k/\sqrt{2}}^{\infty} \exp(-\tfrac{1}{2}t^2)\, dt = \alpha.$$

This is easily available.

The following example shows how to apply the I-procedure.

Consider five normal populations with variances 20, 40, 5, 20 and 10, respectively. Assume a level of 0·05. The critical value is then $k = 2{\cdot}326$. Suppose that the sample sizes are, respectively, 80, 160, 20, 80 and 40, so that $\sigma = 0{\cdot}5$. Let 22·1, 21·4, 23·6, 22·2 and 21·7 be the sample means. Since $u = 2{\cdot}8$, the third population is declared to be the 'best' at a 5 % level.

5. Detection of outliers

In the case of outliers, the least favourable configuration is the null configuration. The critical value k is the solution of the equation

$$W(k, 0, 0, \ldots, 0) = \alpha.$$

The critical values are:

Table 2

no. of populations $n+1$	Level 0·10	0·05	0·01	0·001
3	1·556	1·957	2·738	3·640
4	1·422	1·780	2·489	3·320
5	1·337	1·671	2·339	3·129
6	1·276	1·594	2·235	3·000
7	1·230	1·537	2·158	2·903

The proposed procedure for the detection of outliers is the same as the one defined in §4, but for the new choice of the critical value k. The following example taken from McKay (1935) and Nair (1948) shows how to apply this test for outliers.

In the course of routine testing of a standard leather product of a tannery, five parallel tests yielded the following values for the hide substance content of the leather specimens: 32·44, 36·45, 39·64, 40·13, 41·09. The first observation appears unduly low. Long experience of the product in question has established a value of 2·226 for the standard deviation. The value of u is found to be 1·801, which is significant at the 5 % level.

6. Evaluation of performance

In order to evaluate the performance of the I-procedure, a comparison was made with the M-procedure already described in the preceeding paper (1958).

The critical values for the M-procedure are:

Table 3.

no. of populations $n+1$	Level 0·10	0·05	0·01	0·001
4	1·696	1·941	2·431	3·014
5	1·835	2·080	2·574	3·166

Random numbers from a normal population were taken from Dixon & Massey (1951) and 30,000 tests were carried out in the case of 4 populations, and 20,000 tests with 5 populations. These tests were not independent because four (or five) random numbers were taken and tested after adjustment for the 8 different values of γ. This was then repeated for the remaining 3 (or 4) cyclical permutations. This being done, the first number was discarded and a new number was introduced in the last position and the procedure repeated. Even if these tests were not independent, this does not invalidate the comparison between the two procedures. The following results were obtained.

Table 4. *Four populations*

(*a*) *Probability of decision with the I-procedure*

Level	10%		5%		1%		0·1%	
γ	G	W	G	W	G	W	G	W
0·0	0·030	0·091	0·015	0·044	0·003	0·010	0·000	0·001
0·5	·076	·063	·041	·031	·009	·007	·002	·001
1·0	·154	·041	·092	·020	·026	·004	·005	·000
1·5	·268	·023	·183	·011	·069	·002	·013	·000
2·0	·423	·011	·311	·005	·141	·001	·037	·000
2·5	·585	·004	·469	·002	·252	·000	·086	·000
3·0	·738	·001	·632	·000	·398	·000	·175	·000
3·5	·851	·000	·775	·000	·564	·000	·299	·000

(*b*) *Probability of decision with the M-procedure*

0·0	0·023	0·069	0·011	0·034	0·003	0·008	0·000	0·001
0·5	·058	·048	·031	·023	·008	·005	·001	·001
1·0	·129	·033	·078	·016	·024	·003	·004	·000
1·5	·249	·022	·165	·011	·060	·002	·013	·000
2·0	·406	0·015	0·300	0·008	0·133	0·001	0·036	0·000
2·5	·585	·010	·468	·005	·254	·001	·087	·000
3·0	·747	·006	·645	·003	·414	·000	·181	·000
3·5	·865	·003	·795	·001	·593	·000	·322	·000

G stands for good decision, W stands for wrong decision.

Table 5. *Five populations.*

(*a*) *Probability of decision with the I-procedure*

Level	10%		5%		1%		0·1%	
γ	G	W	G	W	G	W	G	W
0·0	0·022	0·089	0·011	0·046	0·003	0·011	0·000	0·001
0·5	·056	·067	·032	·034	·009	·009	·002	·001
1·0	·131	·051	·081	·027	·024	·005	·005	·000
1·5	·239	·032	·163	·016	·059	·003	·014	·000
2·0	0·389	0·017	0·287	0·008	0·131	0·001	0·038	0·000
2·5	·554	·006	·447	·003	·243	·000	·085	·000
3·0	·708	·002	·610	·001	·388	·000	·173	·000
3·5	·828	·000	·750	·000	·553	·000	·299	·000

(*b*) *Probability of decision with the M-procedure*

0·0	0·023	0·090	0·011	0·046	0·003	0·011	0·000	0·002
0·5	·056	·068	·032	·034	·008	·009	·002	·001
1·0	·118	·046	·075	·025	·024	·006	·005	·000
1·5	·234	·033	·157	·018	·061	·004	·015	·000
2·0	0·394	0·025	0·290	0·014	0·130	0·002	0·039	0·000
2·5	·574	·017	·463	·009	·256	·002	·095	·000
3·0	·736	·011	·641	·006	·420	·001	·189	·000
3·5	·861	·006	·794	·003	·603	·000	·337	·000

An examination of these results shows that no striking difference exists between the performances of these two procedures. There is a slight indication that, for a given level, the probabilities of a good and a wrong decision are a little higher for the *M*-procedure.

References

Dixon, W. J. & Massey, F. J. (1951). *Introduction to Statistical Analysis.* New York: McGraw-Hill Book Company Inc.

Irwin, J. O. (1925). On a criterion for the rejection of outlying observations. *Biometrika*, **17**, 238–50.

McKay, A. T. (1935). The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe. *Biometrika*, **27**, 466–71.

Nair, K. R. (1948). The distribution of the extreme deviate from the sample mean and its studentized form. *Biometrika*, **35**, 118–44.

Zinger, A. & St-Pierre, J. (1958). On the choice of the best amongst three normal populations with known variances. *Biometrika*, **45**, 436–46.

Longest run of consecutive observations having a specified attribute

By E. J. BURR* and GWENDA CANE

University of New England, Armidale, N.S.W., Australia

1. Introduction

Consider an ordered finite sequence of observations which are classified according to possession or non-possession of a specified attribute. The ordering may be temporal, or spatial, or according to any other property measurable on an ordinal scale without ties. In various applications, the attribute may be described by terms such as success, male, defective, positive, greater than the mean, greater than the qth percentile, and so on. In a distribution-free comparison of two samples, after pooling the observations into a single sequence the attribute would be 'belonging to the first sample'. We shall use the terms trial, success and failure to denote an observation, possession and non-possession of the attribute.

In a sequence of n trials resulting in r successes and $s \equiv n-r$ failures, let k_0 be the length of the longest run of consecutive successes observed. On the null hypothesis that the successes and failures occur in random order, k_0 is a statistic with a known probability distribution. There are many alternative hypotheses under which we might expect to obtain significantly large or significantly small values of k_0; for example, local (or temporary) increases in the probability of success, or various types of serial correlation.

Baticle (1935, 1946) and Garwood (1940) have used an asymptotic formula for the distribution of k_0, valid when $s/n \to 0$. Barton & David (1959) propose another asymptotic formula which is always useful in the upper tail, but they remark that it is inaccurate, for moderate values of n, in the mid-range and lower tail.

In this paper, several asymptotic formulae are compared and their domains of usefulness are discussed, with particular emphasis on the lower tail. Alternative formulae are given for the probability that $k_0 = 1$, and a table is given for determining when the values $k_0 = 1, 2$ or 3 are significantly small in all cases in which no suitable asymptotic formula is available.

2. Exact formulae

Let n ordered trials result in r successes and s failures ($0 < r < n, s = n-r$). When these are permuted in all possible ways, the null hypothesis asserts that the $\binom{n}{r}$ permutations are equally probable.

For each integer k in $1 \leqslant k \leqslant r+1$, let $P(k)$ be the probability (when one of these permutations is selected at random) that the longest success run will have length k_0', less than k; that is,

$$P(k) = \Pr\{k_0' < k | n, r\}. \tag{1}$$

Let $t = s+1$. The s failures divide the sequence of successes into precisely t runs (possibly including one or more 'runs of length zero'). Therefore the number of permutations with longest success

* This paper was revised with the partial support of a National Science Foundation Grant at Stanford University.

run $< k$ is equal to the number of ways of selecting in order t integers whose sum is r, each contained in the set $0, 1, 2, \ldots, k-1$; and this equals the coefficient of x^r in the expansion of

$$(1+x+x^2+\ldots+x^{k-1})^t. \tag{2}$$

Hence by multiplying the binomial expansions in $(1-x^k)^t(1-x)^{-t}$ and dividing the coefficient of x^r by $\binom{n}{r}$ we obtain

$$P(k) = \binom{n}{s}^{-1}\left\{\binom{n}{s} - \binom{t}{1}\binom{n-k}{s} + \binom{t}{2}\binom{n-2k}{s} - \ldots\right\}. \tag{3}$$

For each integer $b > 0$ let $a^{(b)} = a(a-1)\ldots(a-b+1)$ if $a \geqslant b$, $a^{(b)} = 0$ if $a < b$. Then (3) may be written in the forms

$$P(k) = 1 - \binom{t}{1}\frac{(n-k)^{(s)}}{n^{(s)}} + \binom{t}{2}\frac{(n-2k)^{(s)}}{n^{(s)}} - \ldots, \tag{4}$$

$$P(k) = 1 - \binom{t}{1}\frac{r^{(k)}}{n^{(k)}} + \binom{t}{2}\frac{r^{(2k)}}{n^{(2k)}} - \binom{t}{3}\frac{r^{(3k)}}{n^{(3k)}} + \ldots, \tag{5}$$

$$P(k) = 1 - \frac{t.r^{(k)}}{1.n^{(k)}}\left[1 - \frac{t-1}{2}\frac{(r-k)^{(k)}}{(n-k)^{(k)}}\left\{1 - \frac{t-2}{3}\frac{(r-2k)^{(k)}}{(n-2k)^{(k)}}(1-\ldots)\right\}\right]. \tag{6}$$

The form (6) is convenient for use with a desk calculator when k is not large. In (5) and (6) it is to be understood that a fraction is zero if its numerator is zero.

Let $N = tk - r - 1$ and $R = N - s$. From the fact that the coefficient of x^r in (2) equals that of x^{kt-t-r} we deduce

$$P(k) = \binom{n}{s}^{-1}\left\{\binom{N}{s} - \binom{t}{1}\binom{N-k}{s} + \binom{t}{2}\binom{N-2k}{s} - \ldots\right\}, \tag{3'}$$

$$P(k) = \frac{N^{(s)}}{n^{(s)}} - \binom{t}{1}\frac{(N-k)^{(s)}}{n^{(s)}} + \binom{t}{2}\frac{(N-2k)^{(s)}}{n^{(s)}} - \ldots, \tag{4'}$$

$$P(k) = \frac{N^{(s)}}{n^{(s)}}\left\{1 - \binom{t}{1}\frac{R^{(k)}}{N^{(k)}} + \binom{t}{2}\frac{R^{(2k)}}{N^{(2k)}} - \ldots\right\}. \tag{5'}$$

We could also have deduced (4′) from the fact that the sum of the series (4) becomes identically zero if it is extended to $t+1$ terms by removing the restriction that $a^{(b)} = 0$ for $a < 0$. The series (3′), (4′), (5′) converge more rapidly than (3), (4), (5) when $N < n$, that is, when $k < 1 + 2r/t$. To test whether an observed value k_0 is significantly small we require the value of $P(k_0+1)$, for which one of (3′), (4′) or (5′) is to be preferred when $r/t \leqslant k_0 < 2r/t$.

In the particular case $k = 2$, $P(2)$ is the probability that $k_0 = 1$, that is, that all successes are isolated. It is used to determine whether an observed value $k_0 = 1$ is significantly small, or whether a value $k_0 = 2$ is significantly large. In this case (2) becomes $(1+x)^t$, so that

$$P(2) = \frac{t!\,(n-r)!}{n!\,(t-r)!} = \frac{t^{(r)}}{n^{(r)}} = \frac{s^{(r-1)}}{n^{(r-1)}}. \tag{7}$$

Other properties and applications of the coefficients in (2) are given by Freund & Pozner (1956).

3. Approximations

If tables of binomial coefficients are not available and k and s (or r in (7)) are large, the terms in formulae (4) to (7) may be computed from

$$\log(a^{(c)}/b^{(c)}) \simeq a'\log a' + (b'-c)\log(b'-c) - b'\log b' - (a'-c)\log(a'-c) + (c/24)\left[(a'^2 - a'c)^{-1} - (b'^2 - b'c)^{-1}\right],$$

where $a' = a + \frac{1}{2}$ and $b' = b + \frac{1}{2}$, and the relative error in $a^{(c)}/b^{(c)}$ is about 1 % at most (when $a = c < b$), but is usually much smaller. In particular, (7) yields

$$P(2) \simeq \frac{(t+\frac{1}{2})^{t+\frac{1}{2}}(t-\frac{1}{2})^{t-\frac{1}{2}}}{(t-r+\frac{1}{2})^{t-r+\frac{1}{2}}(t+r-\frac{1}{2})^{t+r-\frac{1}{2}}} \tag{8}$$

with relative error about 8 % when $r = t$ but less than 3 % when $r < t$. If r/t is small we may use logarithmic power series to show from (8) that

$$P(2) \simeq \exp\left[-\frac{r^{(2)}}{t}\left\{1+\frac{r^2}{6t^2}\right\}\right] \tag{9}$$

with relative error of order $r^6/(15t^5)$.

Barton & David (1959) derive (5) and show that $P(k)$ lies between the sum of the first j terms and the sum of the first $j+1$ terms of the series ($j = 1, 2, 3, \ldots$). They conclude that the formula

$$P(k) = e^{-tv}, \tag{10}$$

where $v = r^{(k)}/n^{(k)}$, yields $P(k)$ with error less than $\frac{1}{2}(tv)^2$, so that (10) is sufficiently accurate when tv is small, that is, in the upper tail of the distribution.

In the lower tail (e.g. $P(k) \leqslant 0{\cdot}05$), we usually have $tv > 2$, the series (3), (4), (5) converge slowly, and the labour involved in a direct summation may be prohibitive. If it happens that $N < n$ we can say at once that $P(k) < N^{(s)}/n^{(s)}$, and if also $tR^{(k)}/N^{(k)}$ is small we have rapid convergence in (3′), (4′) and (5′). Otherwise, we seek to modify the terms of (4) or (5).

Baticle (1935, 1946) and Garwood (1940) modify the terms of (4), for fixed s and large n and k, by putting

$$(n-jk)^{(s)}/n^{(s)} \sim (1-j\alpha)^s \quad (j = 1, 2, \ldots, [r/k]), \tag{11}$$

where $\alpha = k/n$. The relative error in (11) is about $j\alpha s^{(2)}/(2n)$, but this may be reduced by changing α as follows.

Let $n_1 = n - \frac{1}{2}(s-1)$ and $\beta = k/n_1$. Then by using logarithmic power series we obtain

$$(n-jk)^{(s)}/n^{(s)} \simeq (1-j\beta)^s \exp\left[-(s^3-s)j\beta(2-j\beta)/\{24n_1^2(1-j\beta)^2\}\right] \tag{12}$$

with relative error $\sim \{(1-j\beta)^{-4}-1\}s^5/(320n_1^4)$. Hence if α is replaced by β in (11) the relative error there is reduced to about $j\beta s^3/(12n_1^2)$. The terms of (4′) may be treated similarly.

Now let k be small and r, n large, with s/n no longer necessarily small. Let $u = (r/n)^k$. Then $r^{(jk)}/n^{(jk)} \simeq u^j$, so that (5) yields

$$P(k) \simeq (1-u)^t. \tag{13}$$

To make this more precise, it may be shown by using logarithmic power series that

$$r^{(jk)}/n^{(jk)} = u^j \exp\{-\epsilon_1(j^2k^2-jk) - O(\epsilon_2 j^3k^3)\},$$

where $\epsilon_1 = \frac{1}{2}(r^{-1}-n^{-1})$ and $\epsilon_2 = \frac{1}{6}(r^{-2}-n^{-2})$. Hence if $\epsilon_1 j^2k^2$ is small, we have

$$r^{(jk)}/n^{(jk)} \simeq u^j\{1-\epsilon_1(j^2k^2-jk)\} \tag{14}$$

with relative error $\frac{1}{2}\epsilon_1^2 j^4k^4$. Using (14), the series (5) may be summed, giving

$$P(k) \simeq (1-u)^t\{1-\epsilon_1 k^2tu(tu-1)(1-u)^{-2} - \epsilon_1 ktu(1-u)^{-1}\}. \tag{15}$$

There is no simple formula for the error in (15), but for $k \leqslant 4$ it has been found sufficiently accurate (relative error < 10 %), down to $P = 0{\cdot}01$ in the lower tail, for all $r \geqslant 20k$.

In the lower tail, (13) shows that $tu > 1$, and hence (15) shows that $(1-u)^t$ overestimates $P(k)$. Since $e^{-vt} > (1-u)^t$, it follows that (13) is a better approximation than (10) in the lower tail. Near $P = 0{\cdot}01$ we have, from (13), $tu \simeq \log 100$, and (15) shows that the relative error in (13) is about $16\epsilon_1 k^2$. Hence (13) may be used down to the 1 % level in the lower tail if $r > 80k^2(1-r/n)$.

4. Table of critical points

It seems likely that in practical applications, significantly small values of k_0 will usually be 1, 2 or 3. For these k_0, and for certain values of r, the accompanying table gives the critical values of n below which the observed value k_0 is significantly small at the 5 % level or the 1 % level (one-tailed). In order to conserve space, only even values of r have been included. When interpolating to odd values of r, with $k_0 = 1$, a second-difference correction of $-0{\cdot}3$ should be applied when $P = 0{\cdot}05$, and $-0{\cdot}2$ when $P = 0{\cdot}01$. For $k_0 = 2$ or 3, the second-difference correction is negligible. For higher values of r, formulae (9) and (15) are sufficiently accurate at the 5 % and 1 % levels.

Of course, only integral values of n are meaningful, but the critical values are given ‘correct to one decimal place’ in order to make accurate interpolation possible, and also to avoid the question whether the tabulated values fall just inside or just outside the boundary of the critical region.

To illustrate the use of the table, consider a sequence of n trials ($n > 60$) containing 60 successes, and suppose that $k_0 = 2$. Then the table shows that this value of k_0 is significantly small at the 5 % level (one-tailed) if $n \leqslant 252$, and is significantly small at the 1 % level if $n \leqslant 205$.

To demonstrate the adequacy of (15) at the limit of the table, consider the case $n = 205$, $r = 60$, $k_0 = 2$. Here the true value of $P(3)$ from (5) is 0·00976, while (15) yields the approximate value 0·00957, with a relative error of only 2 %. But (13) yields the value 0·0245, which is grossly inaccurate.

Critical points for $k_0 = 1, 2, 3$

For each r, k_0 and P, the value of k_0 is significantly small at the level P (one-tailed) if n is less than the tabulated value. For example, if $r = 38$, the probability that $k_0 \leqslant 3$ is less than 0·05 if $n \leqslant 79$, and less than 0·01 if $n \leqslant 69$.

	$k_0 = 1$		$k_0 = 2$		$k_0 = 3$	
r	$P = 0{\cdot}05$	$P = 0{\cdot}01$	$P = 0{\cdot}05$	$P = 0{\cdot}01$	$P = 0{\cdot}05$	$P = 0{\cdot}01$
4	7·5	—	—	—	—	—
6	15·4	12·3	8·1	—	—	—
8	26·1	19·9	12·6	11·2	—	—
10	39·6	29·3	17·5	15·2	13·2	—
12	55·6	40·4	22·8	19·7	16·9	15·5
14	74·2	53·3	28·6	24·5	20·8	18·9
16	95·6	67·9	34·8	29·6	24·8	22·5
18	119·6	84·2	41·5	35·1	29·0	26·2
20	146·4	102·3	48·5	40·8	33·4	30·1
22	175·7	122·1	55·9	46·8	38·0	34·0
24	207·8	143·6	63·7	53·1	42·7	38·1
26	242·5	166·9	71·8	59·7	47·5	42·3
28	279·9	191·9	80·2	66·6	52·5	46·6
30	319·9	218·7	88·9	73·6	57·6	51·0
32	362·6	247·2	98·0	80·9	62·8	55·5
34	408·1	277·4	107·3	88·5	68·2	60·2
36	456·1	309·4	116·9	96·2	73·6	64·9
38	506·8	343·1	126·9	104·2	79·2	69·7
40	560·2	378·5	137·1	112·4	84·9	74·6
42	616·3	415·7	147·5	120·9	90·7	79·6
44	675·1	454·6	158·2	129·5	96·5	84·6
46	736·5	495·3	169·2	138·3	102·5	89·8
48	800·6	537·7	180·4	147·4	108·6	95·0
50	867·3	581·8	191·8	156·6	114·7	100·3
52	936·8	627·7	203·5	166·0	120·9	105·7
54	1008·9	675·3	215·5	175·6	127·3	111·2
56	1083·6	724·6	227·6	185·4	133·7	116·7
58	1103·6	775·7	240·0	195·3	140·2	122·3
60	1241·2	828·5	252·6	205·5	146·8	127·9
62	—	—	—	—	153·4	133·6
64	—	—	—	—	160·1	139·4
66	—	—	—	—	166·9	145·3
68	—	—	—	—	173·8	151·2
70	—	—	—	—	180·8	157·2
72	—	—	—	—	187·8	163·3
74	—	—	—	—	194·9	169·4
76	—	—	—	—	202·1	175·5
78	—	—	—	—	209·3	181·7
80	—	—	—	—	216·6	188·0

REFERENCES

BARTON, D. E. & DAVID, F. N. (1959). Combinatorial extreme value distributions. *Mathematika*, **6**, 63–76.

BATICLE, E. (1935). Le problème de la répartition. *C.R. Acad. Sci., Paris*, **201**, 862–4.

BATICLE, E. (1946). Le problème des stocks. *C.R. Acad. Sci., Paris*, **222**, 355–7.

FREUND, J. E. & POZNER, A. N. (1956). Some results on restricted occupancy theory. *Ann. Math. Statist.* **27**, 537–40.

GARWOOD, F. (1940). An application of the theory of probability to the operation of vehicular-controlled traffic signals. *J.R. Statist. Soc.*, Suppl. **7**, 65–77.

The moments of the non-central t-distribution*

BY D. HOGBEN, R. S. PINKHAM AND M. B. WILK

The Statistics Center, Rutgers University, New Jersey

1. INTRODUCTION AND SUMMARY

The following note presents a procedure for finding simple explicit expressions for the raw moments of the non-central t. A table of numerical values relevant to the first four central moments and cumulants is given. Application of the table is illustrated in finding the approximate values of the expectation of certain functions of the non-central t.

If X is normally distributed with mean δ and variance 1, and Y^2 is independently distributed as central χ^2 with f degrees of freedom, then the random variable $t = t(\delta, f) = X\sqrt{(f)}/Y$ is said to have a non-central t-distribution. The density function for this random variable is

$$f(t) = \frac{\Gamma(f+1)}{2^{\frac{1}{2}(f-1)}\Gamma(\frac{1}{2}f)\sqrt{(\pi f)}}\left(\frac{f}{f+t^2}\right)^{\frac{1}{2}(f+1)} \exp\left[-\frac{1}{2}\left(\frac{f\delta^2}{f+t^2}\right)\right] Hh_f\left(\frac{-\delta t}{f+t^2}\right),$$

where

$$Hh_f(z) = \int_0^\infty \frac{x^f}{f!} e^{-\frac{1}{2}(x+z)^2} dx.$$

When $\delta = 0$, this reduces to the usual central Student's t-distribution.

Fisher (1931) gave a derivation of the distribution, discussed computational procedures and gave large sample approximations to the variance, γ_1 and γ_2 of the non-central t. Johnson & Welch (1940) presented tables of the non-central t and listed the first three central moments. More extensive tables of the probability integral, the density function and the percentage points of the non-central t for a restricted set of values of δ are given by Resnikoff & Lieberman (1957). Azorin (1952) used the moments of a truncated normal distribution to obtain the frequency function and the moments.

In a recent paper Merrington & Pearson (1958) calculated the first four moments for a considerable number of non-central t-distributions, in order to examine the adequacy of representing the distribution by a Pearson Type IV curve. The formulae which they used were in similar form to those given on p. 3, but no table of the values of the c's as functions of f was available to them.† For the mean t, involving our constant c_{11}, they made direct use, however, of a table of the expectation of $\chi/\sqrt{f}$ (Pearson & Hartley, 1954, Table 35) which up to $f = 100$ has very similar argument entries to our table.

2. DERIVATION OF RAW MOMENTS

Attempts to determine the moments of the non-central t directly using its density may lead to complexities. The moments can, however, be found explicitly by the following indirect procedure.

The joint density for (X, Y^2) is

$$p(x, y^2) = [\Gamma(\tfrac{1}{2}f)\, 2^{\frac{1}{2}(f+1)}\sqrt{\pi}]^{-1} y^{f-2} e^{-\frac{1}{2}(x-\delta)^2 - \frac{1}{2}y^2}.$$

Since p is a density

$$\int_{-\infty}^{\infty}\int_0^{\infty} y^{f-2} e^{-\frac{1}{2}(x^2+y^2)+\delta x} dy^2 dx = [\Gamma(\tfrac{1}{2}f)\, 2^{\frac{1}{2}(f+1)}\sqrt{\pi}]\, e^{\frac{1}{2}\delta^2}.$$

* This work was supported by the Office of Naval Research.

† Prof. Pearson writes that our table would undoubtedly have shortened Mrs Merrington's and his work in computing the third and fourth moments of non-central t.

Continued differentiation of both sides k times with respect to δ gives

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} x^k y^{f-2} e^{-\frac{1}{2}(x^2+y^2)+\delta x}\,dy^2\,dx = [\Gamma(\tfrac{1}{2}f)\,2^{\frac{1}{2}(f+1)}\sqrt{\pi}]\,D^k e^{-\frac{1}{4}\delta^2}.$$

Since $t = X\sqrt{(f)}/Y$, it follows from replacing f by $f-k$ in the above identity and dividing through by the appropriate normalizing constant, that the kth raw moment of the non-central t is

$$E(t^k) = (\tfrac{1}{2}f)^{\frac{1}{2}k}\,\frac{\Gamma\frac{1}{2}(f-k)}{\Gamma(\frac{1}{2}f)}\,e^{-\frac{1}{4}\delta^2} D^k e^{\frac{1}{4}\delta^2}.$$

The function $e^{-\frac{1}{4}\delta^2} D^k e^{\frac{1}{4}\delta^2}$ is similar to the Hermite polynomial $e^{\frac{1}{4}\delta^2} D^k e^{-\frac{1}{4}\delta^2}$, except all coefficients of δ are positive; it satisfies similar recurrence relationships (Fisher, 1931).

3. Central moments and cumulants

The central moments and cumulants are polynomial functions of δ whose coefficients are functions of f. Thus

$$\mu = c_{11}\delta, \quad \mu_2 = c_{22}\delta^2 + c_{20}, \quad \mu_3 = c_{33}\delta^3 + c_{31}\delta,$$

$$\mu_4 = c_{44}\delta^4 + c_{42}\delta^2 + c_{40}, \quad k_4 = \mu_4 - 3\mu_2^2,$$

where

$$c_{11} = \sqrt{(\tfrac{1}{2}f)}\,\Gamma[\tfrac{1}{2}(f-1)]/\Gamma(\tfrac{1}{2}f);$$

$$c_{22} = f/(f-2) - c_{11}^2, \quad c_{20} = f/(f-2);$$

$$c_{33} = c_{11}\left[\frac{f(7-2f)}{(f-2)(f-3)} + 2c_{11}^2\right], \quad c_{31} = \frac{3f}{(f-2)(f-3)}\,c_{11};$$

$$c_{44} = f^2/(f-2)(f-4) - 2f(5-f)\,c_{11}^2/(f-2)(f-3) - 3c_{11}^4;$$

$$c_{42} = 6\{f/(f-2)\}[f/(f-4) - (f-1)\,c_{11}^2/(f-3)];$$

$$c_{40} = 3f^2/(f-2)(f-4).$$

The coefficients c_{ij} are tabulated in Table 1 for 45 values of f from 2 to 1000. The corresponding central moments and cumulants may then be written and computed as polynomials in δ. Table 1 is accurate to as many digits as shown. For intermediate values of f it is recommended that c_{20} and c_{40} be computed directly, whereas linear interpolation on $1/(f-a)$ be used for c_{11}, c_{22}, c_{33}, c_{31}, c_{44} and c_{42}, where a equals 1, 3, 9·8, 4·8, 11·4 and 7·5, respectively.

Additional tables giving central and raw moments and certain other relevant functions with computation and interpolation procedures for selected f between 2 and 1000 and δ from 0 to 10 are given by Hogben, Pinkham & Wilk (1960).

4. Two examples

Moments of functions of random variables may often be well-approximated in terms of Taylor's series involving the central moments of the original variable.

Thus if $Z = g(t)$, then

$$E(Z) \doteq \sum_{i=0}^{k} \frac{\mu_i}{i!}\,\frac{d^i g(t)}{dt^i}\bigg|_{t=\mu},$$

where $\mu_0 = \mu = E(t)$, and $\mu_i = E(t-\mu)^i$, $i > 0$.

Resnikoff & Lieberman (1957) consider the average value of

$$g(t) = \bar{p} = \int_{t/\sqrt{(f+1)}}^{\infty} \frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}x^2}\,dx,$$

where t has a non-central t-distribution with $f = 16$ and $\delta = 8{\cdot}08114$.

Proceeding by direct numerical integration, using their tables, they find $E(\bar{p}) \doteq 0{\cdot}0294$. Proceeding as suggested above, one finds an approximation to $E(\bar{p})$ using the first four central moments of the non-central t-distribution. This value is 0·0295, which is about 1 in 300 greater than Resnikoff & Lieberman's estimate.

As another example, suppose b is distributed as a non-central beta variable with one degree of freedom in the numerator, then

$$b = g(t) = t^2/(f+t^2),$$

where t is the non-central t.

It has been shown by Hogben, Pinkham & Wilk (1961) that

$$E(b) = 1 - \frac{f}{\delta^{f+1}} e^{-\frac{1}{2}\delta^2} \int_0^\delta x^f e^{\frac{1}{2}x^2} dx.$$

For the values $f = 9$ and $\delta = 1$, the correct value of $E(b)$ is 0·1700, while the series approximation with four terms yields an estimate of 0·1730.

Table 1. *Non-centrality coefficients in the moments of the non-central t*

f	c_{11}	c_{22}	c_{20}	c_{33}	c_{31}	c_{44}	c_{42}	c_{40}
2	1·77245	—	—	—	—	—	—	—
3	1·38198	1·090141	3·00000	—	—	—	—	—
4	1·25331	0·429204	2·00000	1·430774	7·51988	—	—	—
5	1·18942	·251956	1·66667	0·391819	2·97354	2·32912	21·7058	25·0000
6	1·15124	·174641	1·50000	·173514	1·72686	0·555627	7·11961	13·5000
7	1·12587	·132419	1·40000	·0958806	1·18216	·220997	3·62848	9·80000
8	1·10778	·106149	1·33333	·0602294	0·886227	·112216	2·25553	8·00000
9	1·09424	·0883494	1·28571	·0411291	·703441	·0658634	1·56996	6·94286
10	1·08372	·0755460	1·25000	·0297802	·580566	·0425458	1·17491	6·25000
11	1·07532	·0659193	1·22222	·0225163	·492853	·0294140	0·924366	5·76190
12	1·06844	·0584307	1·20000	·0175994	·427377	·0213878	·754190	5·40000
13	1·06272	·0524464	1·18182	·0141225	·376782	·0161682	·632497	5·12121
14	1·05788	·0475588	1·16667	·0115764	·336598	·0126046	·541926	4·90000
15	1·05373	·0434943	1·15385	·00965776	·303961	·0100748	·472333	4·72028
16	1·05014	·0400628	1·14286	·00817696	·276960	·00822029	·417451	4·57413
17	1·04700	·0371282	1·13333	·00701076	·254271	·00682400	·373228	4·44615
18	1·04423	·0345905	1·12500	·00607629	·234951	·00574860	·336939	4·33929
19	1·04176	·0323749	1·11765	·00531620	·218311	·00490411	·306697	4·24706
20	1·03956	·0304241	1·11111	·00468978	·203835	·00422971	·281155	4·16667
21	1·03758	·0286935	1·10526	·00416752	·191133	·00368316	·259332	4·09598
22	1·03579	·0271480	1·10000	·00372761	·179900	·00323446	·240493	4·03333
23	1·03416	·0257595	1·09524	·00335365	·169897	·00286183	·224085	3·97744
24	1·03267	·0245054	1·09091	·00303312	160935	·00254919	·209678	3·92727
25	1·03130	·0233672	1·08696	·00275632	·152861	·00228444	·196937	3·88199
30	1·02590	·0189588	1·07143	·00181055	·122131	·00141976	·150530	3·70879
35	1·02209	·0159462	1·06061	·00127910	·101628	·$0^3$964543	·121426	3·59238
40	1·01925	·0137581	1·05263	·$0^3$951236	·0869916	·$0^3$696724	·101573	3·50877
45	1·01706	·0120972	1·04651	·$0^3$734909	·0760262	·$0^3$526289	·0872081	3·44583
50	1·01532	·0107936	1·04167	·$0^3$584744	·0675079	·$0^3$411303	·0763532	3·39674
60	1·01272	·00887919	1·03448	·$0^3$395412	·0551391	·$0^3$270773	·0610691	3·32512
70	1·01088	·00754117	1·02941	·$0^3$285075	·0465944	·$0^3$191529	·0508439	3·27540
75	1·01014	·00701270	1·02740	·$0^3$246471	·0432423	·$0^3$164338	·0469076	3·25584
80	1·00950	·00655342	1·02564	·$0^3$215207	·0403396	·$0^3$142542	·0435333	3·23887
90	1·00843	·00579435	1·02273	·$0^3$168194	·0355638	·$0^3$110177	·0380512	3·21089
100	1·00758	·00519282	1·02041	·$0^3$135056	·0317982	·$0^4$876917	·0337901	3·18878
150	1·00503	·00341823	1·01351	·$0^4$584851	·0207881	·$0^4$369838	·0216437	3·12384
200	1·00377	·00254753	1·01010	·$0^4$324755	·0154402	·$0^4$202676	·0159134	3·09215
300	1·00251	·00168769	1·00671	·$0^4$142490	·0101943	·$0^5$877648	·0104011	3·06095
400	1·00188	·00126180	1·00503	·$0^5$796378	·00760893	·$0^5$487310	·00772428	3·04553
500	1·00150	·00100754	1·00402	·$0^5$507727	·00606957	·$0^5$309461	·00614302	3·03634
600	1·00125	·$0^3$838566	1·00334	·$0^5$351686	·00504825	·$0^5$213793	·00509909	3·03024
700	1·00107	·$0^3$718127	1·00287	·$0^5$257910	·00432113	·$0^5$156492	·00435839	3·02589
800	1·00094	·$0^3$627940	1·00251	·$0^5$197192	·00377709	·$0^5$119482	·00380557	3·02263
900	1·00083	·$0^3$557878	1·00223	·$0^5$155640	·00335473	·$0^6$942025	·00337720	3·02010
1000	1·00075	·$0^3$501880	1·00200	·$0^5$125961	·00301732	·$0^6$761727	·00303550	3·01808

Note: $0^6$761727 should be read 0·000000761727, etc.

References

Azorin, F. (1952). Sobre la distribución *t* no central. *Revista de Real Academie de Ciencias Exactes.* Fisicas y Naturales de Madrid, **46**, 491–5.

Fisher, R. A. (1931). Properties of *Hh* functions. *British Association Mathematical Tables*, **1**, xxxvi. Cambridge University Press.

Hogben, D., Pinkham, R. S. & Wilk, M. B. (1960). The moments of the non-central *t*-distribution. *Technical Report*, no. N-2. The Statistics Center, Rutgers-The State University.

Hogben, D., Pinkham, R. S. & Wilk, M. B. (1961). The moments of a variate related to the non-central *t*. *Technical Report* no. N-4. The Statistics Center, Rutgers-The State University.

Johnson, N. L. & Welch, B. L. (1940). Applications of the non-central *t*-distribution. *Biometrika*, **31**, 362–89.

Merrington, Maxine & Pearson, E. S. (1958). An approximation to the distribution of non-central *t*. *Biometrika*, **45**, 484–91.

Pearson, E. S. & Hartley, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press.

Resnikoff, G. J. & Lieberman, G. J. (1957). *Tables of the Non-Central t-Distribution*. Stanford University Press.

Estimation of missing observations in split-plot experiments where whole-plots are missing or mixed-up

By J. D. BIGGERS

The Wistar Institute, Philadelphia 4, *Pennsylvania*

Summary

The problem of estimating missing split-plots when entire whole-plots are missing is discussed. The simultaneous equations for estimating the missing values are singular and inconsistent, but they can be solved subject to the condition that the sum of the unknowns equals the value which minimizes the whole-plot error mean square. If some of the values are merely mixed-up a direct solution is always available.

1. Introduction

This paper is concerned with estimating missing observations in split-plot experimental designs where some of the missing observations comprise whole-plots. In some types of experiment a common occurrence is loss of a whole-plot. This may occur, for example, with a litter of mice in an experiment where each individual receives a split-plot treatment and each litter a whole-plot treatment. If the animals forming a litter are housed in the same cage some unexpected catastrophe, such as flooding from the water bottle, may destroy them during the experiment. Loss of an entire whole-plot may occur in organ-culture experiments where paired organs from an embryo are used, and the embryo becomes infected while it is being removed from the egg or uterus. The subject of estimating missing plots in split-plot designs where the whole-plots are arranged in randomized blocks and latin squares has been discussed in an earlier paper (Biggers, 1959). There, it was merely stated that when entire whole-plots are missing, the matrix of the equations for estimating the missing values is singular, and there is therefore no simple solution to the problem. Wilkinson (1958) has discussed briefly those cases where the equations are singular and consistent, but has not considered split-plot designs where the estimating equations are both singular and inconsistent.

In the present paper the discussion will apply specifically to split-plot designs where whole-plot treatments are arranged in randomized blocks, since the formulae which are derived can easily be applied to other types of design.

2. Notation

Scalars—letters in italics.
Vectors—lower case letters in Clarendon type.
Matrices—capital letters in Clarendon type.
Transposition of vectors and matrices is indicated by a prime.

3. The problem

Consider a design with r blocks containing t plots to which t whole-plot treatments are allotted at random, and let s split-plot treatments be allotted at random within each whole-plot. We denote the observation made on the kth split-plot treatment in the ijth plot by X_{ijk} ($i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, t$; $k = 1, 2, \ldots, s$). At the end of the experiment let p split-plots be missing, denoted by the corresponding letter in the lower case: x_{ijk}.

The analysis of variance of this design provides two error terms: the error derived from the analysis of whole-plot totals, and the error derived from the analysis of individual observations made on the split-plots. We are concerned with estimating values for the missing observations which minimise the second of these two error terms. It is well known that in this design missing values in a given whole-plot are calculated only from the available observations which have been influenced by the allotted whole-plot treatment. Thus we need only consider the estimation of missing values within a given whole-plot treatment.

Suppose now that in the randomized block design all s observations on the split-plots in the fgth whole-plot are lost. Let I_{gk} be the total of the available observations receiving the gth whole-plot treatment and the kth split-plot treatment, and S_g be the total of the available observations receiving the gth whole-plot treatment. The matrix equation for estimating values for these missing observations which minimize the split-plot error sum of squares is given by

$$\begin{bmatrix} (r-1)(s-1) & -(r-1) & . & . & -(r-1) \\ -(r-1) & (r-1)(s-1) & . & . & -(r-1) \\ . & . & . & . & . \\ . & . & . & . & . \\ -(r-1) & -(r-1) & . & . & (r-1)(s-1) \end{bmatrix} \begin{bmatrix} x_{fg1} \\ x_{fg2} \\ . \\ . \\ x_{fgs} \end{bmatrix} = \begin{bmatrix} sI_{g1} - S_g \\ sI_{g2} - S_g \\ . \\ . \\ sI_{gs} - S_g \end{bmatrix}$$

$$= \begin{bmatrix} (s-1)I_{g1} - \sum_{\substack{k \\ k \neq 1}} I_{gk} \\ (s-1)I_{g2} - \sum_{\substack{k \\ k \neq 2}} I_{gk} \\ . \\ . \\ (s-1)I_{gs} - \sum_{\substack{k \\ k \neq s}} I_{gk} \end{bmatrix} = \begin{bmatrix} (s-1) & -1 & . & . & -1 \\ -1 & (s-1) & . & . & -1 \\ . & . & . & . & . \\ . & . & . & . & . \\ -1 & -1 & . & . & (s-1) \end{bmatrix} \begin{bmatrix} I_{g1} \\ I_{g2} \\ . \\ . \\ I_{gs} \end{bmatrix},$$

since $S_g = \sum_k I_{gk}$. This equation may be written in abbreviated matrix notation

$$\mathbf{Ax} = \mathbf{Ai}/(r-1). \tag{1}$$

The circulant matrix $\mathbf{A}$ in equation (1) is singular and therefore cannot be cancelled from each side of the equation; furthermore, the equations are inconsistent. Thus there is no set of values for the elements of $\mathbf{x}$ which minimizes unconditionally the split-plot error sum of squares.

It is possible, however, to minimize the error sum of squares in the analysis of whole-plot totals using the formulae for estimating missing values in randomized block designs. Let the usual estimate of the missing whole-plot treatment value of the fgth plot be M_{fg}. (In the sequel the subscript will be omitted since no confusion can arise.) Then we can minimize the split-plot error sum of squares subject to the condition that

$$\sum_k x_{fgk} = M_{fg}. \tag{2}$$

4. Some fundamental matrices

We first give some properties of matrices which arise in the sequel. The $d \times d$ square matrix

$$\begin{bmatrix} (s-1)a & -a & . & . & -a \\ -a & (s-1)a & . & . & -a \\ . & . & . & . & . \\ . & . & . & . & . \\ -a & -a & . & . & (s-1)a \end{bmatrix} \tag{3}$$

is a circulant matrix, whose determinant equals $a^d s^{d-1}(s-d)$. Let $\mathbf{C}$ be a matrix of this type of order s.

Then the determinant of $\mathbf{C}$ is zero, and the matrix is singular and of rank $(s-1)$. The adjugate matrix of $\mathbf{C}$ is an $s \times s$ matrix with all its elements equal to $a^{s-1}s^{s-2}$.

Consider now the bordered matrix

$$\mathbf{B} = \left[\begin{array}{c|c} \mathbf{C} & \mathbf{u} \\ \hline \mathbf{u}' & 0 \end{array}\right],$$

where $\mathbf{u}$ is an $s \times 1$ column vector whose elements are either 0 or 1, which can be arranged in any order. Let w of the s elements be 1 and $(s-w)$ of them zero. It can be shown that

$$\det \mathbf{B} = \mathbf{u}' \operatorname{adj} \mathbf{C}\mathbf{u} = -w^2 a^{s-1} s^{s-2}. \tag{4}$$

Thus this matrix is always non-singular, provided $w \neq 0$; $a \neq 0$; $s \neq 0$.

An important special case of (4) is where $w = s$, i.e. all the elements of the vector $\mathbf{u}$ are 1. This matrix is

$$\left[\begin{array}{ccccc|c} (s-1)a & -a & . & . & -a & 1 \\ -a & (s-1)a & . & . & -a & 1 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ -a & -a & . & . & (s-1)a & 1 \\ \hline 1 & 1 & . & . & 1 & 0 \end{array}\right], \tag{5}$$

and its inverse is given by

$$\left[\begin{array}{ccccc|c} (s-1)/(as^2) & -1/(as^2) & . & . & -1/(as^2) & 1/s \\ -1/(as^2) & (s-1)/(as^2) & . & . & -1/(as^2) & 1/s \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ -1/(as^2) & -1/(as^2) & . & . & (s-1)/(as^2) & 1/s \\ \hline 1/s & 1/s & . & . & 1/s & 0 \end{array}\right]. \tag{6}$$

5. Conditional minimization of the split-plot error sum of squares

Equation (2) may be written in matrix notation

$$\mathbf{1}'\mathbf{x} = M, \tag{7}$$

where $\mathbf{1}$ is a unit column vector of length s. We introduce an undetermined multiplier $2\lambda/N$, giving the Lagrange equations

$$\frac{\partial}{\partial \mathbf{x}} \{F(\mathbf{x}) + 2\lambda(\mathbf{1}'\mathbf{x} - M)/N\} = \mathbf{0},$$

where $F(\mathbf{x})$ is the sum of squares for the split-plot error expressed as a function of the vector of missing values. Introducing the notation used in equation (1) of the previous paper (Biggers, 1959) and equation (1) of the present paper, we find that

$$2\{\mathbf{x}'\mathbf{A} - \mathbf{1}'\mathbf{A}/(r-1)\}/N + 2\lambda\mathbf{1}'/N = \mathbf{0}.$$

Transposition and rearrangement gives

$$\mathbf{A}\mathbf{x} + \lambda\mathbf{1} = \mathbf{A}\mathbf{i}/(r-1). \tag{8}$$

The values of x_{fgk} $(k = 1, 2, \ldots, s)$ are given by the simultaneous solution of (7) and (8). These two matrix equations may be combined to give

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{1} \\ \hline \mathbf{1}' & 0 \end{array}\right] \left[\begin{array}{c} \mathbf{x} \\ \hline \lambda \end{array}\right] = 1/(r-1) \left[\begin{array}{c|c} \mathbf{A} & \mathbf{0} \\ \hline \mathbf{0} & (r-1) \end{array}\right] \left[\begin{array}{c} \mathbf{i} \\ \hline M \end{array}\right]. \tag{9}$$

The matrix on the left side of this equation is of the form (5) and will always be non-singular. Thus equation (9) has a unique solution given by

$$\left[\begin{array}{c} \mathbf{x} \\ \hline \lambda \end{array}\right] = 1/(r-1) \left[\begin{array}{c|c} \mathbf{A} & \mathbf{1} \\ \hline \mathbf{1}' & 0 \end{array}\right]^{-1} \left[\begin{array}{c|c} \mathbf{A} & \mathbf{0} \\ \hline \mathbf{0} & (r-1) \end{array}\right] \left[\begin{array}{c} \mathbf{i} \\ \hline M \end{array}\right]. \tag{10}$$

Expansion of (10) with respect to the x_{ghk}'s gives

$$\begin{bmatrix} x_{fg1} \\ x_{fg2} \\ . \\ . \\ x_{fgs} \end{bmatrix} = 1/s(r-1) \left[\begin{array}{ccccc:c} (s-1) & -1 & . & . & -1 & (r-1) \\ -1 & (s-1) & . & . & -1 & (r-1) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ -1 & -1 & . & . & (s-1) & (r-1) \end{array}\right] \left[\begin{array}{c} I_{g1} \\ I_{g2} \\ . \\ . \\ I_{gs} \\ \hdashline M \end{array}\right]. \tag{11}$$

Thus once an estimate of M has been obtained by an analysis of the available whole-plot totals, the values of the missing split-plots can be calculated directly using equation (11). Equation (11) allows us, however, to write formulae for a few simple cases which can be used directly.

(*a*) $s = 2$.

$$x_{fg1} = (I_{g1} - I_{g2} + (r-1)M)/2(r-1),$$
$$x_{fg2} = (I_{g2} - I_{g1} + (r-1)M)/2(r-1).$$

In this case it is usual to avoid the problem just discussed by analysing separately the sums and differences of the split-plots in each whole-plot. Thus in each analysis it is only necessary to estimate values for the missing function using the formula for a randomized block design. Often, however, it is preferable to keep the data in its original form, and use the formulae just given.

(*b*) $s = 3$.

$$x_{fg1} = (2I_{g1} - I_{g2} - I_{g3} + (r-1)M)/3(r-1),$$
$$x_{fg2} = (2I_{g2} - I_{g1} - I_{g3} + (r-1)M)/3(r-1),$$
$$x_{fg3} = (2I_{g3} - I_{g1} - I_{g2} + (r-1)M)/3(r-1).$$

6. The case where some of the values are mixed-up

A more general problem is the case where $(s-w)$ split-plot observations in the fgth whole-plot are missing and the remaining w are mixed-up. Let the sum of the mixed-up values equal H_{fg}. In this case we can always use the method for estimating mixed-up values given previously (Biggers, 1959) and obtain a unique solution for both the mixed-up and the completely missing observations. In this procedure we minimize the error mean square, subject to the condition that the total of the mixed-up values equals H. The matrix is a bordered matrix of the form given by (5), where **u** contains both zero and unit elements. Since this matrix is non-singular in all practical cases, it always has an inverse, and so unique values can be calculated for all the missing observations.

7. More complex cases

Consider an experiment where the fgth whole-plot is entirely missing, and the bcth whole-plot has some of its s split-plots also missing. The estimation of missing values is done in three stages.

(i) Estimate the missing values in the bcth whole-plot using the formula given in the previous paper (Biggers, 1959). The matrix of the estimating equations will be non-singular and so a unique solution is always possible.

(ii) Calculate the total of the bcth whole-plot including the estimates of the missing observations in (i), and use the total in calculating an estimate of the fgth whole-plot total by the formula for the randomized block design.

(iii) Proceed with calculating the conditional estimates of the split-plots in the fgth whole-plot given by formula (11). Extensions to more complex cases are obvious.

Consider now an experiment where the fgth whole-plot is entirely missing and missing observations occur in other whole-plots subjected to the same whole-plot treatment. The matrix of the estimating equations is singular in these cases and in order to solve the equations they will have to be minimized conditionally. However, the matrix will not be of the form (5), and will have to be inverted as a special problem.

8. Reduction of the number of degrees of freedom

The number of degrees of freedom which should be deducted from the number of degrees of freedom for error is determined by the number of constraints placed on the data. The number varies with the problem. In general, the number equals the number of missing or mixed-up observations minus the

number of sets of mixed-up observations. This applies irrespective of whether or not the matrix is non-singular or singular. If there are several submatrices in the complete solution corresponding to different whole-plot treatments, then the total number of degrees of freedom to deduct is the sum of the numbers associated with each submatrix.

References

BIGGERS, J. D. (1959). The estimation of missing and mixed-up observations in several experimental designs. *Biometrika*, **46**, 91–105.

WILKINSON, G. N. (1958). Estimation of missing values for the analysis of incomplete data. *Biometrics*, **14**, 257–86.

Deterministic customer impatience in the queueing system GI/M/1; a correction

By P. D. FINCH

University of Melbourne

1. Introduction

In a recent paper, with the above title, Finch (1960) certain paragraphs were inadvertently left out of the revised version of the manuscript. This has lead to an ellipsis in the argument leading to the fundamental equation (3) of that paper. I am indebted to Mr D. G. Kendall for pointing out the existence of this ellipsis. The argument of the next section should come between equations (1) and (3) of Finch (1960).

2. The correction

Note first that an impatient departure occurs always at the head of the queue. For, if at any time the arrivals at t_n, t_{n+1} are each waiting for service and if w'_n, w'_{n+1} are, respectively, the times they have been waiting then $w'_n = w'_{n+1} + \tau_{n+1}$. Thus the arrival at t_{n+1} cannot depart impatiently and leave the arrival at t_n still waiting for service.

If $w_n < W$ it is easy to see that

$$\left.\begin{aligned} w_{n+1} &= 0 && \text{if} \quad w_n + u_n < 0 \\ &= w_n + u_n && \text{if} \quad 0 \leqslant w_n + u_n < W \\ &= W && \text{if} \quad W \leqslant w_n + u_n. \end{aligned}\right\} \tag{2'}$$

If $w_n = W$, then on the impatient departure of the nth customer some earlier customer is still receiving service, let the remaining service time of this customer be s'_n. Then it is not difficult to show that, if $w_n = W$,

$$\left.\begin{aligned} w_{n+1} &= 0 && \text{if} \quad W + u'_n < 0 \\ &= W + u'_n && \text{if} \quad 0 \leqslant W + u'_n < W \\ &= W && \text{if} \quad W \leqslant W + u'_n \end{aligned}\right\} \tag{2}$$

where $u'_n = s'_n - \tau_{n+1}$. Note that (2)′ is the same as (2) with $w_n = W$ and u_n replaced by u'_n.

Because the service time distribution is exponential, the random variable s'_n has the same exponential distribution as service time and is independent of τ_{n+1}, w_n. This u'_n has the same distribution as u_n. Thus if we define w_n by equation (2) in the case $w_n = W$ also, the distribution of w_n will be that of the waiting time of the nth customer.

Thus we obtain the equation (3) of Finch (1960).

3. The limiting distribution

The sequence of random variables $\{w_n\}$ forms a Markov chain with state space the finite interval $(0, W)$. In Finch (1960) it was stated that it follows from the general theory of Markov chains that a limiting distribution $F(x) = \lim_{n\to\infty} \Pr(w_n \leqslant x)$ exists and is independent of the initial distribution

$\Pr(w_1 \leqslant x)$. Mr D. G. Kendall has pointed out that this limit theorem does not follow from general theory which establishes only that

$$\lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} F_j(x)$$

exists.

A proof of the existence of a limiting distribution for the sequence $\{F_n(x)\}$, where $F_n(x) = \Pr(w_n \leqslant x)$ and the sequence $\{w_n\}$ is given by (2) will be found in Finch (1961).

References

Finch, P. D. (1960). Deterministic customer impatience in the queueing system GI/M/1. *Biometrika*, **47**, 45–52.

Finch, P. D. (1961). A storage problem along a production line of continuous flow. To appear in the *Ann. Univ. Eötvos, Budapest.*

Corrigenda

(1) *Biometrika* (1957), **44**, pp. 399–410

'Further contributions to multivariate confidence bounds'

By S. N. Roy and R. Gnanadesikan

p. 400, equation (2·3) and the following line should read

$$\text{or} \quad \lambda_1 \frac{\mathbf{a}'(T'S_2^{-1}T)\mathbf{a}}{\mathbf{a}'\mathbf{a}} \geqslant \frac{\mathbf{a}'(T'D_{\sqrt{\gamma_i}}S_1^{-1}D_{\sqrt{\gamma_i}}T)\mathbf{a}}{\mathbf{a}'\mathbf{a}} \geqslant \lambda_2 \frac{\mathbf{a}'(T'S_2^{-1}T)\mathbf{a}}{\mathbf{a}'\mathbf{a}}, \tag{2·3}$$

for all non-null $\mathbf{a}(p \times 1)$'s, where $S = TT'$ and T is a lower triangular matrix.

(2) *Biometrika* (1958), **45**, pp. 411–20

'Moment estimators and maximum likelihood'

By L. R. Shenton

p. 412, equation (6) *read* $\frac{\partial \mathscr{A}_j^{(r)}}{\partial \theta_{kr}}$ *for* $\frac{\partial \mathscr{A}_j^{(r)}}{\partial \hat{\theta}_{kr}}$,

line 6 *read* θ *for* $\hat{\theta}$.

lines 10, 12, 14 *read* $\hat{\theta}$ *for* θ.

p. 415, equation (18) *read* $\lambda^x x^x$ *for* $(\lambda\alpha)^x$,

line 21, *read* $\lambda s(s-1)q_{s-1}(x)/\alpha^2$ *for* $\lambda s(s-1)q_{s-1}(x)$,

line 24 *insert* P_x on the right-hand side of equation (22).

p. 419, equation (35) should read

$$\sum_{x=0}^{\infty} \frac{P_x}{x+z} = \frac{1}{z+}\,\frac{\lambda}{1+}\,\frac{1(1+\alpha^{-1}\lambda)}{z+}\,\frac{\lambda(1+\alpha^{-1})}{1+}\,\frac{2(1+\alpha^{-1}\lambda)}{z+}\,\frac{\lambda(1+2\alpha^{-1})}{1+\ldots}.$$

(3) *Biometrika* (1959), **46**, pp. 483–6

'Extrema of quadratic forms with applications to statistics'

By K. A. Bush and I. Olkin

The authors are indebted to D. G. Kendall for pointing out an error in (1′) on p. 485.

(*a*) The paragraph including this formula should read

By making the correspondence $w = x$, $A = M$, $a(B'A^{-1}B)^{-1}B' = y$, we obtain the bound in *I*,

$$wAw' \geqslant \frac{[wB(B'A^{-1}B)^{-1}a']}{a(B'A^{-1}B)^{-1}a'} = a(B'A^{-1}B)^{-1}a',$$

with equality holding if and only if $w = \alpha a(B'A^{-1}B)^{-1}B'A^{-1}$. Because of the constraint $wB = a$, we also have $\alpha = 1$. (Note that $B'A^{-1}B$ is non-singular by conditions I.)

(*b*) p. 486 (ii) should read

(ii) The solution is given directly by (1′) with $A = I$, $a = e$: $1 \times m+1$, and B replaced by B':

$$ww' \geqslant e(BB')^{-1}e',$$

with equality holding if and only if $w = e(BB')^{-1}B$. This is equivalent to the solution of Karush and Wolfsohn, and presents an alternative expression for their result.

K. A. Bush and I. Olkin

(4) *Biometrika* (1960), **47**, pp. 263–71

'Supplemented balance'

By S. C. Pearce

p. 268, six lines from the bottom

read $$\left\{r_0 - \left(\frac{1}{k}\left[s_0 - \frac{r_0\lambda_0}{r}\right]\right)^{+} - \left(\frac{1}{k}\left[s_0 - \frac{r_0\lambda_0}{r}\right]\right)^{0}\right\}\gamma_0$$

in place of the expression given.

(5) *Biometrika* (1960), **47**, pp. 393–8

'On the power function of the exact test for the 2×2 contingency table'

By B. M. Bennett and P. Hsu

p. 395, five lines from the bottom

read $m_2 = \min(r, B)$ *for* $m_2 = \max(r, B)$.

(6) *Biometrika* (1961), **48**, pp. 95–108

'Significant tests for paired-comparison experiments'

By T. H. Starks and H. A. David

p. 97, equation (13) *read* $2^{3n-2nt+1}$ *for* $2^{3n-2n+1}$.

p. 99, equation (26) *should read*

$$f(a_i, a_j, \ldots, a_k) = 2^{-\frac{1}{2}ns(2^t-s-1)} \sum \prod_{p}^{i,j\ldots,k} \binom{n(t-s)}{a_p - a'_p}.$$

p. 102, in last line of footnote to Table 2,

read $\sigma = 0{\cdot}00005\sqrt{(k/3)}$ *for* $\sigma = 0{\cdot}00005(k/3)$.

p. 105, equation (51) *read* $Q_k^2/S_k \leqslant D_\alpha$ *for* $Q_k^2/S_k \leqslant D$.

p. 106, § 8·2. In (3) *read* m'_c *for* m_c.

(7) *Biometrika* (1961), **48**, pp. 115–23

'The use of orthogonal polynomials of positive and negative binomial frequency functions in curve fitting by Aitken's method'

By H. T. Gonin

p. 120, line 4 *read* with mean as origin *for* with zero origin.

(8) *Biometrika* (1961), **48**, pp. 151–65

'Expected values of normal order statistics'

By H. Leon Harter

The author wishes to thank Churchill Eisenhart for calling to his attention the inexcusable failure to mention, in the historical section, the twenty-decimal-place tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution computed by J. Barkley Rosser. A ten-decimal-place version of these tables was published by Teichroew [*Annals of Mathematical Statistics* (1956), **27**, pp. 410–26].

(9) *Biometrika* (1961), **48**, pp. 181–90

'Test of independence in intraclass 2×2 tables'

By M. Okamoto and G. Ishii

p. 181, equation (2) *read* $p_{\overline{AA}} = p_{\overline{A}}^2$ *for* $p_{\overline{AA}} = p_{\overline{A}}$.

p. 184. fourteenth line of § 4 *read* $2np$ *for* $2p$.

sixteenth line of § 4 *read* $N(np^2, np^2q^2)$ *for* $N(2np^2, np^2q^2/4)$.

equation (14) *read* $2n-3$ *for* $2n-s$.

(10) *Biometrika* (1961), **48**, pp. 197–9

'On Durbin's formula for the limiting generalized variance of a sample of consecutive observations from a moving-average process'

By A. M. Walker

p. 197, three lines from the bottom *read* $\sum_{r=1}^{\min(i,j)} b^{ir}b^{rj}$ *for* $\sum_{r=1}^{h} b^{ir}b^{jr}$.

p. 198, line 8 *read* $b^{ij} = \alpha_{j-i}$ *for* $b^{ij} = \alpha_{j-1}$.

line 9 *read* $\alpha_{i-r}\alpha_{j-r}$ *for* $\alpha_{r-i}\alpha_{r-j}$.

line 10 *read* $\sum_{r=0}^{j} \alpha_r\alpha_{r+n-i-j}$ *for* $\sum_{r=0}^{i} \alpha_r\alpha_{r+n-i+j}$.

five lines from the bottom

read $\sum_{t=h+1}^{n} \left(\sum_{i=0}^{h} \beta_i x_{t-i}\right)^2 \Big/ \sigma^2$ *for* $\sum_{t=h+1}^{n} \left(\sum_{i=0}^{h} \beta_i x_{t-i}\right) \Big/ \sigma^2$.

Biometrika (1961), **48**, 3 *and* 4, *p.* 477
Printed in Great Britain

Reviews

An Introduction to Statistical Communication Theory. By DAVID MIDDLETON. New York, Toronto, London: McGraw-Hill Book Company, Inc. 1960. Pp. 1140. £9. 14*s*.

In the preface the author defines statistical communication theory as 'a theory which applies probability concepts and statistical methods to the communication process', and this definition indicates the interest of this book to the modern statistician, who will find most of the more recent developments in statistics, such as decision theory, information theory and stochastic processes, discussed and made use of. The reviewer was somewhat surprised by the author's insistence on calling this work of well over a thousand pages 'an introduction', for the treatment of the various topics is usually very detailed, with an impressive coverage of the relevant literature. However, the author explains that various important topics have been omitted, including linear and non-linear feedback and control systems, coding methods, sequential statistical methods, and certain aspects of noise theory.

The book is divided into four main parts; the first part (six chapters) develops the necessary statistical theory, including that of stationary time-series, noise, and information theory. The second part (five chapters) discusses random noise processes at greater length, dealing in particular with normal processes, and functionals and expansions derived from them. It contains a chapter on the Langevin type of equation, i.e. the direct stochastic equation of the process, and the alternative Fokker–Planck type of equation, which refers to the changing distribution for the process and is a special case of the fundamental 'forward' and 'backward' equations of Kolmogorov.

The third part (six chapters) deals more specifically with special communications systems, such as amplitude and frequency modulation. The statistician will no doubt risk skipping some of these more technical sections, though he should not overlook an exposition of linear prediction and optimum filtering included in this part. The fourth part (six chapters) contains a detailed discussion of signal detection and reception, with a formulation first in general Bayesian terms and then in terms of specific alternative procedures. The problem is of course similar in principle to discrimination procedures generally, and the likelihood ratio criterion is generalized to incorporate prior probabilities and cost functions in order to provide the optimum decision rule. It is noted that two of the particular procedures discussed, the Neyman–Pearson system (in which error 'of the second kind' is minimized for given risk of error 'of the first kind') and the so-called Ideal system (in which both types of possible error are treated symmetrically) can be regarded as special cases of Bayes procedures.

The reviewer occasionally felt in reading this book that he was in danger of not seeing the wood for the trees; while the author does warn us that *a priori* signal probabilities may be hard to obtain, each reader must be prepared to make his own decisions and judgements on the relative value of the various alternative decision procedures discussed.

However, it would seem rather unfair to criticize the book for containing so much, and while most statisticians may think twice before buying a book at this price, they can be assured that there is plenty in this volume to engage their attention if they are able to procure a copy.

M. S. BARTLETT

Introduction to Linear Programming. By WALTER W. GARVIN. New York, Toronto and London: McGraw-Hill Book Company, Inc. 1960. Pp. 281. 68*s*.

This book is based on a course of lectures given to scientists, engineers and economists of the Standard Oil Company of California, to teach them how to use linear programming, and in particular how to recognize and formulate linear programming problems. The book should be equally valuable to people interested in other fields of application, who, without necessarily being professional mathematicians, want to understand the mathematical basis of the subject—which is in fact fairly elementary.

Part I (80 pp.) discusses the general linear programming problem. It has five chapters entitled Introduction, The Simplex Method, The Computational Procedure, Sensitivity Analysis, and A Gasoline

Blending Problem. Part II (65 pp.) discusses the transportation problem and its variants. It has five chapters entitled The Transportation Problem, Unbalance and Trans-shipment, Assignment Problems and the Caterer Problem, A Tanker-routing Problem, and the Generalized Transportation Problem. Part III (118 pp.) entitled Special Methods, has eight chapters entitled Upper Bounds, Statistical Linear Programming, The Revised Simplex Method, Resolution of Degeneracy, Parametric Linear Programming, A Simple Economic Model, Duality, and the Warehouse Problem. Finally, there are four pages of reference and a seven-page index.

The author is concerned exclusively in this book with Dantzig's Simplex Method and modifications of it. This seems entirely reasonable, but I think he could usefully have included a chapter on Lemke's Dual Method. This deserves to be justified in its own right, and not merely as the Simplex Method applied to the dual problem, both because of its fundamental role in Integer Programming and other types of non-linear programming, and—perhaps more relevantly—because of its value in what the author calls 'Sensitivity Analysis', i.e. the study of the effects on the solution of changes or additions to the assumptions of the problem. Indeed, Orchard-Hays's parametric linear programming procedure, described in Chapter 15, can most easily be thought of as using the dual method in the way that Gass and Saaty use the original simplex method to solve for a parametric objective function; although historically it was developed independently.

The chapter on the warehouse problem is included to describe an application of the duality principle. I did not enjoy it because, as A. J. Hoffman pointed out in a private communication, the problem can be solved much more simply by transforming it into a special case of the purchase-storage problem discussed by Beale, Morton and Land, 'Solution of a Purchase–Storage Programme' *Operational Research Quarterly*, Vol. 9, pp. 174–97 (1958). (Each time period is subdivided into a selling period in which requirements not filled from the warehouse are imagined to be satisfied by purchase at the selling price, followed by a purchase period with zero requirements.)

The other weak chapter is that on statistical linear programming. This field includes several problems. The constraints may be known but some coefficients in the objective function may be random variables. These coefficients may be replaced by their mean value if one is content to optimize the mean value of the objective function, but one is faced with a (standard type of) quadratic programming if one wants to minimize the variance of the objective function for a given mean. This problem has been studied by H. M. Markowitz in connexion with investment portfolio selection. Other problems arise if some coefficients in the constraints are regarded as random variables. One may be interested in the resulting distribution of the optimum value of the objective function. But often the key decisions have to be taken before all the values of the unknowns are known. Some of the variables may then have to be selected knowing only the distributions of the random elements, and one may ask how this can be done to optimize the mean value of the objective function, assuming the other variables are appropriately chosen. Without going into these distinctions, the author discusses a problem of this last type, which is made linear by approximating the distributions of the random variables (actually the right-hand sides) by discrete distributions.

In spite of these blemishes, the book is a good introduction to linear programming. For the sake of readers with a limited mathematical background, the author avoids matrix notation. This does not prevent him from giving a detailed and lucid account of the 'Revised Simplex' or 'Inverse Matrix' method of organizing computations (though he does not discuss the product form of the inverse used in most large-scale computer programs for linear programming).

Mathematics are not avoided, but are accompanied by a non-technical commentary on what is going on, and why. There are also many numerical examples, and problems for the reader after each chapter.

E. M. L. BEALE

Applied Statistical Decision Theory. By Howard Raiffa and Robert Schlaifer. Harvard Business School. London: Bailey Bros. and Swinfer Ltd. 1961. Pp. xxviii + 356. 76*s*.

This book is addressed to persons who desire to employ statistical methods in practical problems of decision making. Although the approach is from first principles with regard to decision theory, it is essential that the reader be quite familiar with statistical methods and concepts in addition to possessing a sound mathematical background.

The work is in three parts: general theory, analysis when sampling and terminal utilities are additive, and distribution theory. It reads with the dryness of a handbook. This does not detract from

its value; as a handbook it is excellent and the reviewer has no hesitation in recommending it as a work of reference for those who are already reasonably conversant with the commercial problems of decision making. On the other hand it is felt that the book will not be of immediate value to those looking at the subject for the first time. Decision theory is a very interesting branch of applied statistics and presents some interesting problems for the theorist. It is perhaps a pity that the authors did not consider it necessary to introduce a series of examples and applications since this might well have increased the book's area of appeal.

J. G. SAW

Strategy and Market Structure. Competition, Oligopoly, and the Theory of Games. By MARTIN SHUBIK. Foreword by OSKAR MORGENSTERN. London: Chapman and Hall; New York: John Wiley and Sons. 1959. Pp. 387. 64*s*.

The traditional problems of economics are the allocation of resources and the distribution of income. These problems require a theory of relative prices (of factors as well as final products). The great triumph of nineteenth-century economics was to show that profit maximization by many independent producers in atomistic competition yielded a system of resource allocation and income distribution which depended directly upon the 'real' situation of the economy—tastes, factor supplies, and the technologically given possibilities of production. This solution could be 'distorted' by monopoly; but, while monopoly presented economists with a normative problem, the positive problem presented no theoretical difficulty because cost and demand functions could still be taken as given by the real forces of the economy (the former by technical possibilities and factor prices, the latter by consumers' tastes and incomes). Consider, however, the case of a small number of firms in close competition with each other. Simple maximization of the difference between two given functions will not do: the price and output (and quality and advertising rates) of each firm affect the others, whose reactions must be allowed for in the formulation of policy. Now profit maximization ceases to be a simple concept, and relative prices, allocation, and income distribution depend upon market structure and the policy of firms towards their rivals as well as upon real forces.

We have faced this oligopoly problem for a long time without making much of it, but since the publication of von Neumann and Morgenstern's *Theory of Games and Economic Behaviour* we have been told repeatedly that the solution lies in game theory. Up till now, there has been plenty of propaganda but not much pay-off. Now a most distinguished game theoretician, Prof. Shubik, has made a full-dress onslaught on the problem in his *Strategy and Market Structure*; and it is obviously important to assess its success. This, I confess, I find remarkably difficult. The book is able, powerful, even exciting, but seems to leave two fundamental questions unanswered: (*a*) (assuming that we are still interested in allocation and income distribution) what propositions does the theory yield on these matters? (*b*) how might the theory be tested? At times it seems that the answer to (*a*) is 'nothing' and to (*b*) 'never' because there are so many 'ifs' floating around (as Prof. Shubik recognizes, p. 226); but this is only an impression, and it is to be hoped that it is a wrong one. Prof. Shubik does not, in fact, address himself to these questions—it would probably require a second volume to do this, and it is to be hoped that he will write it. In the meantime one can only guess about the answers.

What Prof. Shubik has done in this volume is to offer, first, a critique of existing oligopoly theory from the point of view of game theory and, secondly, his own dynamic theory of oligopoly, the 'game of economic survival' (plus two chapters of 'application'). The first part serves to introduce the reader to game theory and to show how most existing theories of oligopoly can be made to 'fall out' as particular cases of game theory. My only serious criticism of this part (apart from having to plod through Cournot, Bertrand *et al.* once again, inescapable I suppose, for heuristic reasons, but tiresome none the less) is that Shubik, in spite of his opening remarks about the desirability of testable content, overlooks the well-known method of obtaining testable predictions from static theory by differentiating the solution values of the variables with respect to the parameters.

The second part, in which Prof. Shubik presents his games of economic survival, is the most arresting part of the book. This treatment is extremely powerful, and enormously extends the range of behaviour and initial conditions that can be handled systematically. The asset structure of the firm appears in the theoretical structure, instead of merely in the long list of things that economic theory 'ought to' comprehend (this list is the stock-in-trade of any institutionalist critic of current theory: the point is that Prof. Shubik can relate market behaviour to factors like asset structure and the structure of control and ownership in a systematic, analytic way). And this part is full of illuminating ideas, testimony to the power of the analysis, of which the following are examples.

(1) (p. 257). The lower the interest rate, the greater the incentive for the strong firm to start a 'war of ruin' in order to secure a monopoly position (because, in weighing the prospective profits from the exploitation of the monopoly against the current losses of the war, the former need not be so heavily discounted).

(2) (Chapter 11). The stability of a market equilibrium depends on the power of players to make plausible threats of 'police action' against potential disturbers of the equilibrium, and may be defined in terms of the number of firms who would have to co-operate to make the threat effective.

(3) (p. 243). A firm's 'vulnerability' is defined as the derivative of its profits with respect to a strategy change by the rival. The 'damage exchange rate' is the ratio of the change in the rival's profits to the change in profit of the firm which adopts a 'damage' strategy. (If the strategy is a price cut which the rival firm must follow, the 'damage exchange rate' necessarily favours the smaller firm which loses revenue on fewer units of output. This is the consequence of defining the rate in absolute terms; it might be better to define it in elasticity terms, i.e. as the ratio of the percentage loss of profit by firm i to the percentage loss by firm j for j's change in strategy. In any case, its importance to the plausibility of threats is obvious.)

The book ends with a discussion of the American automobile and tobacco industries in terms of the game theory of Part II, and some reflexions on the Anti-Trust Laws. Both chapters are interesting and illuminating; but is the whole apparatus of game theory necessary to justify the conclusion that General Motors is the dominant firm in the automobile industry, could afford competitive price cuts that would ruin all rivals save Ford, but dare not because of the Sherman Act? Is it necessary to justify the conclusion that the major cigarette firms are so much of-a-muchness that it wouldn't pay any of them to start a 'war to the ruin'? (We get the good point here that major firms may be able to get rid of minor ones without warfare, merely by accepting a stable price and leaving random fluctuations to take care of the tiddlers.)

It only remains to say that *Strategy and Market Structure* is, in the opinion of this reviewer, an important and very able book. It may not yet justify all the claims that have been made for game theory, but it is an essential preliminary to testing them.

(There are a few minor lapses that might be removed in a second edition. The Frisch–Allen concept of 'conjectural variation', analysed in the latter's *Mathematical Analysis for Economists*, is worth explicit mention. In Chapter VI.2 the matter of imputed rent appears to have been overlooked. Throughout that chapter the unusual assumption is made that entry is a function of price instead of profit. In Fig. 9, p. 76, not all the curves are labelled. Fig. 1, p. 15, which is supposed to elucidate an elementary point in game theory, remains a mystery to me. On pp. 148–9 a kinked oligopoly demand curve is described as dynamic: it is merely a special case of conjectural variation. The English is often heavy going—words like 'effectivity' and 'interconnectivity' are not necessary.) G. C. ARCHIBALD

Regression Analysis. By R. L. PLACKETT. Oxford: Clarendon Press and Oxford University Press. 1960. Pp. 173. 35*s*.

The book discusses the various aspects of Least Square theory in connexion with its application to the analysis of experimental data. A thorough understanding of the content demands a familiarity with the theory of matrices and complex integration, probably at the level of the senior mathematics undergraduate, and with these mathematical tools Plackett develops the theory clearly and concisely. As well as the anticipated topics (linear hypotheses, polynomial regression, stationary error processes, symmetrical factorial experiments) the book contains other information such as methods for inverting a matrix and the effect of departure from standard test conditions. If the entirely adequate manner in which the more familiar theory is discussed were not enough to recommend it, such sections as those on the condition for the independence of two quadratic forms in normal variables or the power of the variance-ratio test, taken in conjunction with the interesting examples and list of references at the end of each chapter would serve to make *Regression Analysis* a valuable addition to the bookshelf.

J. G. SAW

Probability. An Introduction. By SAMUEL GOLDBERG. U.S.A. and London: Prentice Hall, Inc. 1960. Pp. 322. 54*s*.

If the student, reading this book without the added help of a lecture course, finds the going heavy then it is because of the subject-matter itself and not the fault of the author who discusses each topic with care and clarity and often with a relieving humour.

Throughout the book, populations are held to be finite so that the differential and integral calculus is never called into use. Work begins with a discussion of the definition and algebras of sets and advances from these foundations to definitions and methods of manipulating probabilities including conditional and joint probabilities; the mean and variance of a population are also defined and discussed. In the last part of the book, the author discusses and develops the theory of Bernoulli trials and the Binomial distribution and introduces the idea of testing a hypothesis statistically by referring to the standard 'quality control' techniques when sampling for defectives. The worked examples are plentiful and entertaining and the exercises seem carefully designed to test and increase the readers appreciation of each section. Answers and hints to aid the solution of half these exercises are given at the end of the book.

Probability is perhaps ideally suited to the undergraduate in mathematics as a companion to a course on the applications of set theory. The statistics student has to concern himself with a series of increasingly sophisticated applications of probability theory, however, and it is perhaps quite common nowadays to invoke his almost intuitive concept of probability and to proceed from there. If this be the case, then such a student may well be recommended to read this book at the end of his first year of statistics, first to gain an appreciation of the logical argument and secondly to find firm ground for the future development of the subject.

J. G. SAW

Elementary Statistics. By PAUL G. HOEL. New York and London: John Wiley and Sons, Inc. 1960. Pp. 261. 44*s*.

The author writes in the preface that his book is 'designed for a one-semester course for students whose background in mathematics is limited to high-school algebra'. Consequently this book will come as a shock to many since it would seem that the high-school graduate is unequipped to draw the graph $y = 2+3x$ and needs to be told that 'the line rises three units vertically for every positive horizontal unit change'.

Though we are introduced to sampling, theoretical frequency distributions, testing hypotheses, correlation and regression it is not surprising that since a page of detail needs be devoted to the description and drawing of a line there is little hope of a mature discussion on the principle of least squares and its application to line fitting. The last four chapters on the chi-square distribution, non-parametric tests, analysis of variance and time series are written in similar vein as special topics. Each chapter concludes with a set of examples.

The book is very poor in references to either classical or contemporary authors in statistics; this is a pity because, although 'Elementary Statistics' is intended to be a descriptive work, it should at least make some provision somewhere for the reader who is sufficiently stimulated to ask the question 'why'.

J. G. SAW

Sampling Methods for Censuses and Surveys. By F. YATES. Third edition, revised and enlarged. 1960. Pp. xvi+440. 54*s*.

This is the third edition of a book which first appeared in 1949. It appears to be a reprint of the second edition (1953) with the addition of Chapter 11 on the use of electronic computers in the analysis of censuses and surveys. In this last chapter the general principles governing the working of electronic computers and their programming are set out and useful instructions are given regarding their application to problems considered previously.

F. N. DAVID

Statistische Methoden der Populationsgenetik. By HENRI LOUIS LE ROY. Basel and Stuttgart: Birkhauser Verlag. 1960. Pp. 397. 67.50. Sw. Fr.

To the English or American reader, the title of this book may be somewhat misleading as it does not deal with the subject developed by Fisher, Haldane and Sewall-Wright and thus is not comparable with Li's (1955) book *Population Genetics*. Rather it covers a few aspects of quantitative genetics, the analysis of variance as a means for estimating the effects of genetical and environmental factors and the application of this knowledge to the selective practices of breeders. It thus provides German-speaking readers with an introduction and some elaborations of the methods of Lush, Kempthorne and to some extent of Mather, Lerner and Falconer. As such it appears quite excellent. Starting in a fairly elementary manner it leads on to many detailed descriptions of more complex situations illustrating them by some numerical working examples. Sixty-five line drawings provide valuable help for visualizing the logical structure of many of the calculations. Another unusual feature of the book is the substitution at the end of three tables of content for a conventional index. In addition to a list of references used there is a large additional bibliography covering a wider field than the text.

The book is beautifully produced.

H. KALMUS

Symposia of the Society for Experimental Biology. No. XIV. Models and Analogues in Biology. Ed. J. W. L. BEAMENT. Cambridge University Press. 1960. Pp. 255. 50*s*.

The eighteen contributions in this volume are exceedingly diverse in content—though there is not a single botanical contribution—and to the reviewer's mind also vary in relevance. A couple of quite excellent specialist papers could have been published elsewhere without raising any suspicion that they had anything to do with models or analogues; the pertinence of an introductory paper by the most venerable father of atomic theory is also problematical, while the editor states in the general conclusions to his own contribution that 'Practical biophysical models can be particularly useful in the investigation of the properties of non-living organic materials and systems'. The other authors are more to the point, and very refreshingly they approach their tasks from quite different directions. Only a few attempt elaborate definitions of what they mean by a model or by an analogue, though most use the terms to include mathematical or logical hypotheses and imaginative propositions, as well as—sometimes rather grudgingly—a few devices, which a small boy might recognize as models.

The following branches of biology are considered: genetics, development, tissue culture, muscle physics, animal locomotion, flight, behaviour, neurology and biology teaching. And the models and analogues described are derived from statistics, kinetics, electrical engineering, cybernetics, biochemistry and biophysics as well as from other biological specialities. Readers of this journal might be professionally interested in the contribution of J. H. Westcott on the 'estimation of values of parameters of a model to conform with observations'.

The volume is well produced but perhaps a little expensive.

H. KALMUS

Numerical Methods of Curve Fitting. By P. G. GUEST. London: Cambridge University Press. 1961. Pp. xii + 422. 80*s*.

This is a book written by a physicist primarily for other physicists. The title is perhaps a little misleading. This is essentially a book on the fitting of regression curves and there is, for instance, scarcely any mention of the fitting of frequency functions or of time series analysis.

The book is in three parts:

Part I (82 pp.). Distribution theory, tests of hypotheses, confidence limits, etc.

Part II (64 pp.). Straight line regression.

Part III (263 pp.). Polynomial and other regression.

It is the third part which will be of most value since the first two parts are more fully covered in many of the standard introductions to statistical theory. The third part includes, for instance, the fitting of power series, of orthogonal polynomials when the observations are equally spaced including the case where some of the values are missing, the use of grouped observations, of harmonic analysis and of multiple regression.

For someone who has frequently to fit curves to sets of data this should prove an excellent handbook. Every method is described in detail and accompanied by examples, calculating schemes and in many places by tables of coefficients which the author has obviously found useful in his own work. Derivations are given of almost all of the results required. The author does not use matrix notation but since much of the literature on regression uses it, there is a section devoted to an explanation of the use of matrices.

ALAN J. MILLER

Characteristic Functions. By E. LUKACS. Griffin's Statistical Monographs and Courses, no. 5—Double Volume. London: C. Griffin and Co. 1960. Pp. 216. 38*s*.

This book is an account, from the standpoint of pure mathematics, of the theory of characteristic functions of probability distributions on the real line. The first four chapters treat the basic theory. The rudiments of distribution function theory are first reviewed without proofs. The characteristic function and the moments are then defined, and their elementary properties established. The uniqueness theorem is proved, without inversion, by an approximation method, and then follows an elementary (i.e. without dominated convergence) proof of the inversion formula. The convolution formula and the continuity theorem are proved. Then simple necessary conditions are found for a function to be a characteristic function, and Bochner's characterization (i.e. 'Bochner's theorem') is proved together with three other less well-known characterizations. These four chapters also treat the most familiar distributions as examples and discuss several special classes of distributions (lattice, absolutely continuous, singular, etc.).

The author does not consider applications of the theory, and he excludes the central limit theorem and results of that type (so fully treated by B. V. Gnedenko & A. N. Kolmogorov: *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Cambridge, Mass., 1954). The rest of the book (more than half) is instead devoted almost entirely to factorization problems.

Chapter V treats the elementary properties of infinitely divisible distributions and then the canonical representations of Lévy–Khintchine, Lévy, and Kolmogorov. The standard examples are cited, and stable distributions briefly considered. The next chapter treats the general factorization problem. More detailed results are obtained in the succeeding two chapters devoted to analytic characteristic functions (i.e. those admitting analytic extension to a horizontal strip in the complex plane). The analytic behaviour of the characteristic function is related to properties of the distribution. Necessary conditions for analytic functions to be characteristic functions are found (including results of the author and the theorem of Marcinkiewicz) by application of the theory of entire functions. A remarkable result of Zinger and Linnik about analytic characteristic function solutions of differential equations is proved. The well known theorems of Cramér and Raikov on factorization of normal and Poisson distributions are proved. Other problems are considered in these chapters, with many examples. There is a brief final chapter on integral transforms of distribution functions.

The book is a comprehensive and well-organized account of a subject in which the modern sources are widely scattered, and will become a standard reference for all users of characteristic function theory. Prospective buyers should be warned that the first printing contained many minor errors; the text has now been reset.

D. A. EDWARDS

The Elliptic Functions as they should be. By ALBERT EAGLE. Cambridge: Galloway and Porter. 1958. Pp. 508. 45*s*.

Nearly all statisticians require the properties of elliptic functions at some time or other in their career and most of these have to resort hastily to a text-book to refresh their memories or to learn anew. The difficulty has been until now to find a book which will so supplement one's mathematical education without burdening oneself unduly with techniques irrelevant to one's main purpose. Mr Eagle has written an excellent treatise which contains all that the user of applicable mathematics will want and in which he delineates the theory in an easily comprehensive way.

In the book he covers twenty elliptic functions and their forty trigonometric series, derivatives and squares of elliptic functions, theta and peeta functions, the addition theorems, elliptic integrals of the

first, second, third and fourth kinds, the kay and eee functions, the weierstrassian poleless functions, applications of the elliptic functions in geometry and mechanics, conformal representation and the elliptic functions in it. There are examples mostly of the 'prove that' variety with a short line where the author thinks the proof may cause difficulty. The reviewer was not helped, however, by the remark at the end of the statement of a problem that this 'is extraordinarily easy'.

This book can be unreservedly recommended to all students and research workers seeking for knowlege of the elliptic functions. The author has certain eccentricities of nomenclature and notation (such, for example, as writing $\frac{1}{2}\pi$ as τ and calling it 'hi') but these are easily learnt and should not put anyone off learning from him.

F. N. DAVID

Finite-difference Methods for Partial Differential Equations. By GEORGE E. FORSYTHE and WOLFGANG R. WATSON. London and New York: John Wiley and Sons, Inc. 1960. Pp. 444. 92*s*. or $11.50.

When a partial differential equation is replaced by an approximating finite difference equation, an error is introduced, the estimation of which is as formidable a task for a mathematician as the solution of the difference equation is for even a high speed electronic computer. The accumulation of rounding off errors is also very difficult to control. The authors seem to have explained this difficult theory as lucidly as possible.

The reader need not know anything about partial differential equations but should know enough analysis for an honours degree in mathematics; and some experience of using electronic computers on easier problems would help. A reader not so qualified will find that Lebesgue integration is used but not defined; and that the introductory section on computers assumes without explicitly saying so that they use the binary scale.

G. L. WATSON

The Calculus of Finite Differences. By GEORGE BOOLE. Fourth edition. New York: Chelsea Publishing Co. 1960. Pp. vii + 336. $1.39 (paper), $3.95 (cloth).

Boole's *Finite Differences* was written in 1860 and has stood the test of time. It is still in the reviewer's opinion the best text-book from which to learn this branch of mathematical statistics and probability, and the Chelsea Publishing Company are to be congratulated on reprinting it as a paper-back. It may be recalled that the work covers interpolation, quadrature, the summation of series, Bernouilli numbers and factorial powers, difference-equations of the first order, linear difference-equations with constant and with variable coefficients, mixed and partial difference-equations. Exercises and answers are given. The price makes it possible for all students to possess this standard work.

F. N. DAVID

Fundamentals of College Algebra. By W. H. DURFEE. New York and London: The Macmillan Co, New York. 1960. Pp. xi + 250. 31*s*. 6*d*.

This is a text-book of classical algebra presented from the modern point of view. The topics covered are sets, numbers, equations, complex numbers, polynomials, unequalities, groups, exponents and logarithms. It will be of little interest to would-be students of statistics—with exception perhaps of the last chapter on the approximate solution of equations—since the topics discussed are not advanced far enough. Exercises and solutions are given.

F. N. DAVID

Other books received

Statistical Theory and Methodology in Science and Engineering. By K. A. Brownlee. London and New York: John Wiley and Sons. 1960. Pp. xv + 570. 134*s.* or $16.75.

Mathematical Thinking in the Measurement of Behavior. By James S. Coleman, Ernest W. Adams and Herbert Solomon (Ed.). Illinois: The Free Press of Glencoe. 1961. Pp. 314. $7.50.

Developments in Mathematical Psychology. By R. Duncan Luce (Ed.), Robert R. Bush and J. C. R. Licklider. Illinois: The Free Press of Glencoe. 1960. Pp. 294. $7.50.

Field Computations in Engineering and Physics. By A. Thom and Colin J. Apelt. Foreword by G. F. J. Temple. London: D. Van Nostrand Co. 1961. Pp. 165. 30*s.*

General Systems. Yearbook of the Society for General Systems Research. Vol. V. Michigan: Society for General Systems Research. 1960. Pp. 235. $7.50.

Sample Design in Business Research. By W. Edwards Deming. London and New York: John Wiley & Sons. 1960. Pp. 517. 96*s.*

Projective Geometry of *n* Dimensions. By Otto Schreider and Emanuel Sperner. New York: Chelsea Publishing Co. 1961. Pp. 208. $4.95.

Textbook of Algebra. Reprints of Vols. I and II. 6th edition. By G. Chrystal. New York: Chelsea Publishing Co. 1959. Vol. I, 571 pp. Vol. II, 616 pp. $2.95 each.

Quantitative Methods in Pharmacology. Proceedings of a Symposium held in Leyden, May 1960. Ed. H. de Jonge. Amsterdam: North-Holland Publishing Co. 1961. Pp. xx + 391. 84*s.*

Arithmetic: An Introduction to Mathematics. By L. Clark Lay. New York and London: The Macmillan Company. 1961. Pp. 323. 31*s.* 6*d.*

Decomposition of Superpositions of Distribution Functions. By Pál Medgyessy. Budapest: Publishing House of the Hungarian Academy of Science. 1961. Pp. 227.

Lebruch der Praktischen Statistik. By Peter Quante. Berlin: Walter de Gruyter and Company. 1961. Pp. 443. DM30.

Analysing Qualitative Data. By A. E. Maxwell. London: Methuen & Co; New York: John Wiley and Sons. 1961. Pp. 163. 16*s.*

L'Anglais Economique Dictionnaire de Concepts. Paris: Éditions Cujas. 1961. Pp. 199.

International Work in Health Statistics, 1948–1958. By H. S. Gear, Y. Biraud and S. Swaroop. Geneva: World Health Organization. 1961. Pp. 56. 3*s.* 6*d.*, $0.60 or Sw. Fr. 2.

A First Course in Mathematical Statistics. By G. E. Weatherburn (Reprinted in paper-back edition). Cambridge: The University Press. 1961. Pp. 277. 17*s.* 6*d.*

Probability. A First Course. By Frederick Mosteller, Robert E. K. Rourke and George B. Thomas. Massachusetts and London: Addison-Wesley Publishing Company. 1961. Pp. 319. 30*s.*

Probability with Statistical Applications. By Frederick Mosteller, Robert E. K. Rourke and George B. Thomas. Massachusetts and London: Addison-Wesley Publishing Company. 1961. Pp. 478. 38*s.*

Tables of the Hypergeometric Probability Distribution. Stanford Studies in Mathematics and Statistics. III. By Gerald J. Lieberman and Donald B. Owen. Stanford: University Press; Oxford: University Press. 1961. Pp. 726. £6.

Markov Learning Models for Multiperson Interactions. By Patrick Suppes and Richard C. Atkinson. Stanford: University Press; Oxford: University Press. 1961. Pp. 296. 66*s.*

General Notes drawn up for the Guidance of Authors on the Preparation of Manuscripts for *Biometrika*

Papers should be headed by an informative title, by the name or names of the author(s), and where appropriate by the name and address of the laboratory or institution where the work was performed. Manuscripts should normally be typed in double spacing with wide margins and on one side of the paper only; they should be on sheets of uniform size, numbered consecutively. In addition to the top copy it is editorially convenient, though not essential, that a second copy of the paper be submitted to the journal. It is helpful if a shortened version of the title, not exceeding 65 letters and spaces, or about eight words in length, could be supplied for a page-head.

Special care should be taken with mathematics, and any complicated mathematical expressions should be inserted very clearly in the manuscript. Where algebraic expressions are written in, particular attention should be given to distinguishing between capitals and lower-case letters, e.g. S, s, etc. Where Greek letters or unusual symbols are used it may help the printer if they are distinguished by writing or underlining in coloured ink, and a key provided for the compositor. Alternatively each unusual letter may be indicated by name in the margin at the first mention, e.g. 'μ = Greek mu', 'R_F = italic capital R and sub-script italic capital F'. Letters in script should also be indicated in the margin. Equation numbers should be placed after, not before, the equation.

Tables, except very small ones, should be on separate sheets of paper, since they are not set up at the same time as the body of the text; the position at which they are to be inserted should be indicated on the manuscript. They should have headings or footnotes which make their general meaning comprehensible to the reader without reference to the text.

Illustrations should be kept to a minimum. The author's name and the number of the figure should be written on the back of each, with legends on a separate sheet. Diagrams that are not drawn by professional draughtsmen generally have to be re-drawn and re-lettered before blocks are made, and for this reason it is advisable to draw illustrations carefully in pencil from which the printer will make good ink drawings. It is most important that no mistake should be made in any part of the author's original drawing.

The Harvard system is used for references in this journal, references being collected at the end of the paper in alphabetical order of author. Each reference must give name followed by initials of author, the year of publication in brackets, title of paper, journal title abbreviated in accordance with the *World List of Scientific Periodicals*, volume number in arabic numerals with wavy line to indicate black type and the number of the first and last page in arabic numerals.

Papers should be submitted in their final form so that proof correction is confined to a minimum.

BIOMETRIKA PUBLICATIONS

Issued by the Cambridge University Press, Bentley House, London, N.W. 1
and obtainable from any bookseller

Tables of the Incomplete B-Function — EDITED BY KARL PEARSON

59 pages of Introduction and 494 pages of Tables — *Price: 70s. net*

Tables of the Incomplete Γ-Function — EDITED BY KARL PEARSON

31 pages of Introduction and 164 pages of Tables — *Price: 42s. net*

Tables of the Complete and Incomplete Elliptic Integrals

(from LEGENDRE'S *Traité des Fonctions Elliptiques*. With autographed portrait of LEGENDRE)

39 pages of Introduction by KARL PEARSON and 94 pages of Tables — *Price: 12s. 6d. net*

Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples — BY F. N. DAVID

38 pages of Introduction, 55 pages of Tables, 10 Diagrams and 4 Charts — *Price: 17s. 6d. net*

Biometrika Tables for Statisticians, Vol. I

EDITED BY E. S. PEARSON and H. O. HARTLEY for the Biometrika Trust

102 pages of Introduction and 136 pages of Tables — *Price: 25s. net*

The Life, Letters and Labours of Francis Galton, Vols. I, II, IIIA, & IIIB

BY KARL PEARSON, F.R.S. — *Price: £5. 5s. net*

Karl Pearson: An Appreciation of Some Aspects of his Life and Work

BY E. S. PEARSON — *Price: 15s. net*

A Bibliography of the Statistical and Other Writings of Karl Pearson

COMPILED BY G. M. MORANT, with the assistance of B. L. WELCH — *Price: 6s. net*

"Student's" Collected Papers — EDITED BY E. S. PEARSON and JOHN WISHART with a FOREWORD BY LAUNCE MCMULLEN — *Price: 21s. net*

Karl Pearson's Early Statistical Papers

Reprinted by photo-lithography for the Biometrika Trust, with the permission of the original publishers. The Volume contains eleven papers, including the more important of the memoirs entitled "Mathematical Contributions to the Theory of Evolution", first published in the *Philosophical Transactions* of the Royal Society. The original paper deriving the χ^2-distribution, published in 1900 in the *Philosophical Magazine*, is also included. *Price: 25s. net*

PUBLICATIONS OF THE DEPARTMENT OF STATISTICS, UNIVERSITY COLLEGE, LONDON

Issued by the Cambridge University Press, Bentley House, London, N.W. 1
and obtainable from any bookseller

TRACTS FOR COMPUTERS

I. **Tables of the Digamma and Trigamma Functions.** By ELEANOR PAIRMAN, M.A.
Tables for summing $S = \sum_{i=1}^{\infty} \frac{1}{(p_1 i+q_1)(p_2 i+q_2)\ldots(p_n i+q_n)}$ where the p's and q's are numerical factors. *Price 7s. 6d. net.*

V. **Table of Coefficients of Everett's Central-Difference Interpolation Formula.** By A. J. THOMPSON, PH.D. *Second edition. Price 7s. 6d. net.*

VIII. **Table of the Logarithms of the Complete Γ-Function (to ten decimal places) for Argument 2 to 1200 beyond Legendre's Range (Argument 1 to 2).** By EGON S. PEARSON, D.SC. *Price 7s. 6d. net.*

IX. **Log Γ (x) from $x = 1$ to 50·9 by intervals of 0·01.** By JOHN BROWNLEE, M.D., D.SC. *Price 7s. 6d. net.*

X. **On Quadrature and Cubature or on Methods of Determining Approximately Single and Double Integrals.** By J. O. IRWIN, D.SC. *Price 7s. 6d. net.*

XII. **Tables of the Probable Error of the Coefficient of Correlation.** By KARL HOLZINGER, PH.D. *Price 7s. 6d. net.*

XIII. **Bibliotheca Tabularum Mathematicarum,** being a Descriptive Catalogue of Mathematical Tables. Part I. A, Logarithms of Numbers. By JAMES HENDERSON, PH.D. *Price 9s. net.*

XV. **Random Sampling Numbers.** By L. H. C. TIPPETT, M.SC., with a Foreword by KARL PEARSON. *Price 7s. 6d. net.*

XXIII. **Tables of $\tan^{-1}x$ and $\log(1+x^2)$.** To assist in the calculation of the ordinates of a Pearson Type IV curve. By L. J. COMRIE, PH.D. *Price 7s. 6d. net.*

XXIV. **Random Sampling Numbers (2nd Series).** By M. G. KENDALL and B. BABINGTON SMITH. *Price 7s. 6d. net.*

XXV. **Random Normal Deviates.** By HERMAN WOLD. *Price 7s. 6d. net.*

XXVI. **Correlated Random Normal Deviates.** By E. C. FIELLER, T. LEWIS and E. S. PEARSON. *Price 10s. 6d. net.*

Nos. II, III, IV, VI and VII are out of print

LOGARITHMETICA BRITANNICA

A standard Table of Logarithms to Twenty Decimal Places. By A. J. THOMPSON, Ph.D. (commenced in 1922 to commemorate the tercentenary of the publication of HENRY BRIGGS'S *Arithmetica Logarithmica*).

The nine separate sections of this Table have now been issued, and the complete work consisting of the logarithms of numbers 10,000–100,000, together with Dr Thompson's General Introduction (98 pp.), is available in two bound volumes.

Price £8. 8*s.* 0*d.*

NEW STATISTICAL TABLES: SEPARATES RE-ISSUED FROM BIOMETRIKA

To be obtained from

BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON, W.C.1

1. From *Biometrika*, Vols. 22, 27 and 28

Tests of Normality. By E. S. PEARSON and R. C. GEARY — Price 2s. 6d., post free

2. From *Biometrika*, Vol. 32, pp. 168–181 and 188–189

(1) Table of percentage points of the incomplete beta-function

(2) Table of percentage points of the χ^2 distribution

Stitched together with introductory matter. Price 2s. 6d., post free

3. From *Biometrika*, Vol. 32, pp. 300–310

(1) Table of the probability integral of the range in samples from a normal population

(2) Table of the percentage points of the range

(3) Table of the percentage points of the t-distribution

Stitched together with introductory matter. Price 2s. 6d., post free

4. From *Biometrika*, Vol. 33, pp. 73–88

Table of percentage points of the inverted beta (F) distribution

With introductory matter. Price 2s. 6d., post free

5. From *Biometrika*, Vol. 33, pp. 252–265

(1) Table of the probability integral of the mean deviation in samples from a normal population

(2) Table of the percentage points of the mean deviation

Stitched together with introductory matter. Price 2s. 6d., post free

6. From *Biometrika*, Vol. 33, pp. 296–304

Table for testing the homogeneity of a set of estimated variances

With introductory matter. Price 2s., post free

7. From *Biometrika*, Vol. 35, pp. 145–156

Table of significance levels for the Fisher-Yates test of significance in 2 x 2 contingency tables. By D. J. FINNEY — With introductory matter. Price 2s. 6d., post free

8. From *Biometrika*, Vol. 35, pp. 191–201

Table for the calculation of working probits and weights in probit analysis. By D. J. FINNEY and W. L. STEVENS — With introductory matter. Price 2s. 6d., post free

9. From *Biometrika*, Vol. 36, pp. 267–289

Tables of autoregressive series. By M. G. KENDALL — With introductory matter. Price 2s. 6d., post free

10, 14, 18, and 20. From *Biometrika*, Part 1 from Vol. 36, pp. 431–449, Parts 2 and 3 from Vol. 38, pp. 435–462, Part 4 from Vol. 40, pp. 427–446, and Part 5 from Vol. 42, pp. 223–242

Tables of symmetric functions. By F. N. DAVID and M. G. KENDALL

With introductory matter. Price 14s. 6d., post free

(Part 1, 2s. 6d.; Parts 2 and 3, 4s.; Part 4, 4s.; Part 5, 4s.)

11. From *Biometrika*, Vol. 39, p. 190 and Vol. 43, pp. 449–451

Tables of percentage points of the extreme "Studentized" deviate from the sample mean. By K. R. NAIR and H. A. DAVID — With introductory matter. Price 1s., post free

12. From *Biometrika*, Vol. 37, pp. 168–172 and pp. 313–325

(1) Table of the probability integral of the t-distribution

(2) Table of the χ^2 integral, and of the cumulative Poisson distribution. By H. O. HARTLEY and E. S. PEARSON — Stitched together with introductory matter. Price 5s., post free

NEW STATISTICAL TABLES: continued

13. From *Biometrika*, Vol. 38, pp. 112–130

Charts of the power function for analysis of variance tests, derived from the non-central F-distribution. By E. S. PEARSON and H. O. HARTLEY
With introductory matter. Price 2s. 6d., post free

15. From *Biometrika*, Vol. 38, pp. 423–426

A chart for the incomplete beta-function and the cumulative binomial distribution. By H. O. HARTLEY and E. R. FITCH With introductory matter and ruler scale. Price 2s. 6d., post free

16. From *Biometrika*, Vol. 40, pp. 70–73

Tables of the angular transformation. By W. L. STEVENS
With introductory matter. Price 1s., post free

17. From *Biometrika*, Vol. 40, pp. 74–86

Tests of significance in a 2 x 2 contingency table: extension of Finney's table (No. VII). Computed by R. LATSCHA With introductory matter. Price 2s. 6d., post free

19. From *Biometrika*, Vol. 41, pp. 253–260

Tables of generalized *k*-statistics. By S. H. ABDEL-ATY With introductory matter. Price 2s., post free

21. From *Biometrika*, Vol. 42, pp. 494–511

A new form of table for significance tests in a 2 x 2 contingency table. By P. ARMSEN
With introductory matter. Price 2s. 6d., post free

22. From *Biometrika*, Vol. 43, pp. 388–403

Tables for certain applications of sequential methods in the analysis of variance. By W. D. RAY
With introductory matter. Price 2s. 6d., post free

23. From *Biometrika*, Vol. 43, pp. 423–435

Table for determining confidence limits for a proportion in binomial sampling. By EDWIN L. CROW With introductory matter. Price 2s. 6d., post fre

24. From *Biometrika*, Vol. 44, pp. 411–419

Tables for estimating the normal distribution function of normit analysis. Part I. Tables and description of their use. By JOSEPH BERKSON With introductory matter. Price 2s. 6d., post free

25. From *Biometrika*, Vol. 44, pp. 482–489

Table of significance points for a two-sample t-test based on range. By P. G. MOORE
With introductory matter. Price 2s. 6d., post free

26. From *Biometrika*, Vols. 44 & 45

Tables of the upper percentage points of the generalized beta distribution. By F. G. FOSTER and D. H. REES With introductory matter. Price 5s., post free

27. From *Biometrika*, Vol. 46, pp. 178–204

Tables of 1000 standardized random deviates from certain non-normal distributions. By M. H. QUENOUILLE. With a note by E. S. Pearson With introductory matter. Price 2s. 6d., post free

28. From *Biometrika*, Vol. 46, pp. 441–453

Table of confidence limits for the expectation of a Poisson variable. By EDWIN L. CROW and ROBERT S. GARDNER With introductory matter. Price 2s. 6d., post free

29. From *Biometrika*, Vol. 47, pp. 121–142

Nomograms for fitting the logistic function by maximum likelihood. By JOSEPH BERKSON
With introductory matter. Price 5s., post free

30. From *Biometrika*, Vols. 47 & 48

Tables of Neyman-shortest unbiased confidence intervals: (*a*) for the binomial parameter; (*b*) for the Poisson parameter. By COLIN R. BLYTH and DAVID W. HUTCHINSON
With introductory matter. Price 2s. 6d. post free

Please send remittance, made payable to *Biometrika*, crossed 'a/c BIOMETRIKA TRUST' to the SECRETARY, BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, GOWER STREET, LONDON, W.C.1, with order. To pay for an order in dollars, add 2s. for bank charges to the total sum due in sterling before converting to dollars. [*E.g.* For one copy each of Tables VII, XVII and XXI pay 7s. 6d. sterling, but 9s. 6d. = $1.33 (@ $2.80 to £1) if paying by dollar cheque.]

SEPARATES RE-ISSUED FROM BIOMETRIKA

To be obtained from

BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON, W.C.1

The Biometrika Office has available for sale copies of many of the papers which have been published in *Biometrika* (right back to Volume 1). Readers requiring offprints of any papers are invited to enquire from the above address whether they are still available. The following are a few examples of fairly recent separates of which copies may be obtained:

From *Biometrika*, Vol. 39, pp. 324–345

Rank Analysis of Incomplete Block Designs I. The Method of Paired Comparisons. By R. A. BRADLEY and M. E. TERRY Price 4s., post free

From *Biometrika*, Vol. 41, pp. 502–537

Rank Analysis of Incomplete Block Designs II. Additional Tables of Paired Comparisons. By R. A. BRADLEY Price 5s., post free

From *Biometrika*, Vol. 43, pp. 203–205

Further critical values for the two-means problem. By W. H. TRICKETT, B. L. WELCH and G. S. JAMES Price 1s., post free

From *Biometrika*, Vol. 44, pp. 1–8

John Wishart, 1898–1956. Obituary Notice and Bibliography. By E. S. PEARSON Price 2s., post free

From *Biometrika*, Vol. 44, pp. 490–514

A bibliography on the theory of queues. By ALISON DOIG Price 5s., post free

From *Biometrika*, Vol. 45, pp. 293–315

THOMAS BAYES'S Essay towards solving a problem in the doctrine of chances. [Reproduced from *Phil. Trans. Roy. Soc.* 1763, 53, 370–418.] With a biographical note by G. A. BARNARD Price 5s., post free

From *Biometrika*, Vol. 45, pp. 521–543

A bibliography on life testing and related topics. By WILLIAM MENDENHALL Price 5s., post free

BIOMETRIKA INDEX. Comprising Subject Index for Vols. 1–37 and Author Index for Vols. 1–40, with Author Index Supplement covering Vols. 41–43. Price 6s. or $1.00, post free

STATISTICAL EXERCISES

Issued by the DEPARTMENT OF STATISTICS
UNIVERSITY COLLEGE LONDON

Elementary Statistical Exercises. By F. N. DAVID and E. S. PEARSON

Published by the Cambridge University Press

These exercises were collected in connection with the numerical classwork undertaken by students during the first year of the B.Sc. Special Degree course in Statistics at University College. They deal with applications of the more elementary univariate and bivariate theory.

Answers are provided. *Price* 13*s.* 6*d.*

Statistical Exercises (Part II). Analysis of variance and associated techniques.

Compiled by N. L. JOHNSON

This volume includes over 100 exercises on the following topics:

Analysis of variance techniques for the following experimental designs: randomized block, Latin square, confounding, split plots, fractional replication. *Linear and curvilinear regression, Dosage-mortality techniques, Discriminant functions, Time series, Curve fitting.* *Price* 12*s.* 6*d.*

This part is at present reproduced by offset-litho processes from typescript and bound in stiff paper covers. Copies may be obtained from a bookseller or direct from the Department of Statistics, University College, Gower Street, London, W.C. 1. If ordered direct, payment must be made in advance and cheques made payable to University College, London.

[A supplement containing suggested solutions to a substantial proportion of the exercises in Part II is in preparation.]

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOL. 56, NO. 294 CONTENTS JUNE 1961

ALLAN BIRNBAUM—Confidence Curves: An Omnibus Technique for Estimation and Testing Statistical Hypotheses.
EDWIN B. COX—Change in the Size Distribution of Dividend Income.
WALTER D. FISHER—Note on Curve Fitting with Minimum Deviations by Linear Programming.
E. J. GUMBEL—Bivariate Logistic Distributions
ROBERT L. GUSTAFSON—Partial Correlations in Regression Computations.
CAROL M. JAEGER and JEAN L. PENNOCK—An Analysis of Consistency of Response in Household Surveys.
DALE W. JORGENSON—Multiple Regression Analysis of a Poisson Process.
CLYDE Y. KRAMER and LEROY S. BRENNE—Factorial Treatments in Rectangular Lattice Designs.
H. O. LANCASTER—Significance Tests in Discrete Distributions.
SHIRLEY Y. LEHMAN—Exact and Approximate Distributions for the Wilcoxon Statistic with Ties.
FRED C. LEONE, C. W. TOPP, and Y. H. RUTENBERG—The Use of Sample Quasi-Ranges in Setting Confidence Intervals for the Population Standard Deviation.
M. N. MURTHY and V. K. SETHI—Randomized Rounded-off Multipliers in Sampling Theory.
D. S. ROBSON and CHITRA VITHAYASAI—Unbiased Componentwise Ratio Estimation.
EUGENE ROGOT—A Note on Measurement Errors and Detecting Real Differences.
LOWELL E. GALLAWAY and PAUL E. SMITH—A Quarterly Econometric Model of the United States.
R. F. TATE—On the Use of Partially Ordered Observations in Measuring the Support for a Complete Order.
GERHARD TINTNER—The Statistical Work of Oskar Anderson.
COLERIDGE A. WILKINS—A Problem Concerned with Weighting of Distributions.

VOL. 56, NO. 295 CONTENTS SEPTEMBER 1961

IRMA ADELMAN and ZVI GRILICHES—On an Index of Quality Change.
EDWARD AMES and STANLEY REITER—Distributions of Correlation Coefficients in Economic Time Series.
F. J. ANSCOMBE—Estimating a Mixed-Exponential Response Law.
R. L. BASMANN—A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in two Leading Over-Identified Cases.
JOSEPH BERKSON—The Other Side of the Lower Bound. A Note with a correction.
MURRAY BROWN—*Ex Ante* and *Ex Post* Data in Inventory Investment.
JACK CAPON—A Note on the Asymptotic Normality of the Mann-Whitney-Wilcoxon Statistic.
M. DAVIES—Multiple Linear Regression Analysis with Adjustment for Class Differences.
CAROL B. EDWARDS and JOHN GURLAND—A Class of Distributions Applicable to Accidents.
P. R. FISK—Estimation of Location and Scale Parameters in a Truncated Grouped Sech Square Distribution.
MAX HALPERIN—Fitting of Straight Lines and Prediction when both variables are Subject to Error.
H. LEON HARTER—The Use of Sample Ranges in Setting Exact Confidence Bounds for the Standard Deviation of a Rectangular Population.
BERNARD LAZERWITZ—A Comparison of Major United States Religious Groups.
G. R. LINDSEY—The Progress of the Score during a Baseball Game.
J. NEYMAN and E. L. SCOTT—Further Comments on the 'Final Report of the Advisory Committee on Weather Control.'
PAUL PEACH—Bias in Pseudo-Random Numbers.
JOHN W. PRATT—Length of Confidence Intervals.
S. M. SHAH—A Note on Griffin's paper 'Graphic Computation of Tau as a Coefficient of Disarray'.
RALPH THOMLINSON—A Model for Migration Analysis.

For further information, please contact:

AMERICAN STATISTICAL ASSOCIATION

BEACON BUILDING, 1757 K STREET, N.W., WASHINGTON 6, D.C.

ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

VOL. 29, NO. 2 CONTENTS APRIL 1961

HERBERT A. SIMON and ALBERT ANDO: Aggregation of Variables in Dynamic Systems.

FRANKLIN M. FISHER: On the Cost of Approximate Specification in Simultaneous Equation Estimation.

PETER R. FISK: The Graduation of Income Distributions.

PATRICK SUPPES: Behavioristic Foundations of Utility.

H. O. HARTLEY: Nonlinear Programming by the Simplex Method.

M. MORISHIMA and F. SETON: Aggregation in Leontief Matrices and the Labour Theory of Value.

LEIF JOHANSEN: A Note on 'Aggregation in Leontief Matrices and the Labour Theory of Value'.

A. L. NAGAR: A Note on the Residual Variance Estimation in Simultaneous Equations.

YASUSUKE MURAKAMI: A Note on the General Possibility Theorem of the Social Welfare Function.

LIONEL W. MCKENZIE: On the Existence of General Equilibrium: Some Corrections.

REPORT OF THE ST LOUIS MEETING.

REPORT OF THE STANFORD MEETING.

BOOK REVIEWS.

LETTER TO THE EDITOR.

ANNOUNCEMENTS AND NOTES.

VOL. 29, NO. 3 CONTENTS JULY 1961

L. R. KLEIN: A Model of Japanese Economic Growth, 1878–1937.

MICHAEL C. LOVELL: Manufacturers' Inventories, Sales Expectations and the Acceleration Principle.

JOHN F. MUTH: Rational Expectations and the Theory of Price Movement.

LESLIE KISH: Efficient Allocation of a Multi-Purpose Sample.

FREDERICK V. WAUGH: The Place of Least Squares in Econometrics.

YEHUDA GRUNFELD: The Interpretation of Cross Section Estimates in a Dynamic Model.

RAGNAR FRISCH: A Reconsideration of Domar's Theory of Economic Growth.

DENIS SARGAN: The Maximum Likelihood Estimation of Economic Relationships with Auto-Regressive Residuals.

PETER E. DE JANOSI: The Statistical Discrepancy in the National Accounts Revisited.

E. MALINVAUD: The Estimation of Distributed Lags: A Comment.

REPORT OF THE STANFORD MEETING.

REPORT OF THE ST LOUIS MEETING.

BOOK REVIEWS.

ANNOUNCEMENTS AND NOTES.

RICHARD RUGGLES, *Secretary*

THE ECONOMETRIC SOCIETY, BOX 1264, YALE UNIVERSITY

NEW HAVEN, CONNECTICUT, U.S.A.

TRABAJOS DE ESTADISTICA

REVIEW PUBLISHED BY *INSTITUTO DE INVESTIGACIONES ESTADÍSTICAS*
OF THE *CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS*
MADRID-6, SPAIN

Vol. XII CONTENTS *Cuaderno I and II*

H. BRENY—Un théorème de la théorie statistique des modèles lineaires.

M. DWASS—The distribution of linear combinations of random divisions of an interval.

M. M. RAO—Some remarks on independence of statistics.

S. RIOS, J. BEJAR and J. M. GARCIA—Un problema de normalización de mididas como base para un problema de producción e inventario.

J. C. SMIT—Estimation of the mean of a stationary stochastic process by equidistant observation.

S. RIOS—Procesos de decision.

NOTAS.

S. RIOS—La investigación operativa y los métodos clásicos de organización.

J. M. GARCIA—La investigación operativa en la industria química.

I. YAÑEZ—Nota sobre programación lineal estocástica.

CRONICA BIBLIOGRAFIA CUESTIONES PROPUESTAS EJERCICIOS

For everything in connection with works, exchanges and subscriptions write to Professor Sixto Ríos, Instituto de Investigaciones Estadísticas, C.S.I.C., Serrano 123, Madrid-6, Spain.

The Review is composed of three fascicles published three times a year (about 350 pp.), and its annual price is 100 pesetas for Spain and 240 pesetas for all other countries.

BIOMETRICS

Journal of the Biometric Society

Vol. 17, No. 3 CONTENTS September 1961

The Growth and Age-Distribution of a Population of Insects Under Uniform Conditions. E. J. WILLIAMS—Empirical Sampling Estimates of Genetic Correlations. L. D. VAN VLECK and C. R. HENDERSON—A New Method of Testing Hypotheses and Estimating Parameters for the Logistic Model. JAMES E. GRIZZLE—A Generalized Model of a Host-Pathogen System. C. J. MODE—Latin Squares to Balance Immediate Residual, and Other Order, Effects. PAUL R. SHEEHE and IRWIN D. J. BROSS—On the Statistical Theory of a Roving Creel Census of Fishermen. D. S. ROBSON—Allocation of Experimental Units in Some Elementary Balanced Designs. J. S. WILLIAMS—Augmented Designs with One-Way Elimination of Heterogeneity. WALTER T. FEDERER—Phenotypic, Genetic and Environmental Correlations. S. R. SEARLE—Relative Efficiencies of Heritability Estimates Based on Regression of Offspring on Parent. B. B. BOHREN H. E. MCKEAN and YUKIO YAMADA—Note: On a Formula for the Estimation of the Optimum Dressing of a Fertilizer. F. PIMENTAL GOMES—Book Reviews: S. K. Ekambaram: The Statistical Basis of Quality Control Charts. L. R. SHENTON—Gerald J. Lieberman and Donald B. Owen: Tables of the Hypergeometric Probability Distribution. R. A. BRADLEY—K. C. S. Pillai: Statistical Tables for Tests of Multivariate Hypotheses. M. J. R. HEALY—F. Yates: Sampling Methods for Censuses and Surveys. M. D. MOUNTFORD.

Vol. 17, No. 4 CONTENTS December 1961

The Poisson Pascal Distribution. S. K. KATTI and JOHN GURLAND—Some Rank Sum Multiple Comparison Tests. ROBERT G. D. STEEL—The Estimation of Repeatability and Heretability from Records Subject to Culling. R. N. CURNOW—Three Classes of Univariate Discrete Distributions. C. G. KHATRI and I. R. PATEL—Further Consideration of Methodology in Studies of Pain Relief. PAUL MEIER and SPENCER M. FREE, Jr.—Fitting a Geometric Progression to Frequencies. E. J. WILLIAMS—Computing Procedures for Estimating Components of Variance in the Two-way Classification. S. R. SEARLE and C. R. HENDERSON—Optimum Sample Size in Animal Disease Control. A. W. NORDSKOG, H. T. DAVID and H. B. EISENBERG—Numerical Aspects of the Regression of Offspring on Parent. H. E. MCKEAN and B. B. BOHREN—Some Hypotheses Concerning Two Phase Regression Lines. P. SPRENT—Note: A Practical Application of a Theoretically Inefficient Design. C. P. COX—Query: On a Graphical Sequential Test. I. D. J. BROSS—Note: A Simple Method of Fitting the Regression Curve. B. K. SHAH—Query: On a Follow-up Study of the Growth of Children. M. R. SAMPFORD—Book Review: H. De Jonge: Quantitative Methods in Pharmacology. D. J. FINNEY.

'Biometrics' is published quarterly. Its objects are to describe and exemplify the use of mathematical and statistical methods in biological and related sciences, in a orm assimilable by experimenters. The annual non-member subscription rate is $7.00. *Inquiries, orders for back issues and non-member subscriptions should be addressed to:*

OFFICE OF *BIOMETRICS*
DEPARTMENT OF STATISTICS, FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA, U.S.A.

Annals of Human Genetics

Formerly ANNALS OF EUGENICS

Edited by L. S. PENROSE

Vol. 25, Pt. II CONTENTS *November 1961*

Frequency of colour blindness in Andhra Pradesh school children. K. R. DRONAMRAJU and P. MEERAKHAN.

Mongoloid twins with 48 chromosomes ($AA+A_{21}XXY$). T. W. J. HUSTINX, P. EBERLE, S. J. GEERTS, J. TEN BRINK, and L. M. WOLTRING.

A factorial analysis of sex-ratio data. A. W. F. EDWARDS.

Rare genetic conditions among the Caingang Indians. F. M. SALZANO.

Stillbirths and infant mortality in twins. A. BARR and A. C. STEVENSON.

Hand clasping (Spaniish data). J. PONS.

A male with *XXYY* chromosomes. J. R. ELLIS, O. J. MILLER, L. S. PENROSE and G. E. B. SCOTT.

Chromosomal mosaicism in a case of Klinefelter's syndrome associated with thalassaemia. M. D'A. CRAWFURD

Enlarged satellites and multiple malformations in the same pedigree. J. R. ELLIS and L. S. PENROSE.

Fibrocystic disease of the pancreas: A comment. M. G. BULMER.

REVIEWS.

Subscription price £5 net per volume of four parts. Single issues, 30s. (postage extra)

CAMBRIDGE UNIVERSITY PRESS

BENTLEY HOUSE, 200 EUSTON ROAD, LONDON, N.W. 1

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

EDITED BY P. C. MAHALANOBIS

Series A, Vol. 23, Part 1

Foreword. P. C. MAHALANOBIS—Sampling the reference set. SIR RONALD A. FISHER—Simple approximations to the probability integral and $P(\chi^2, 1)$ when both are small. J. B. S. HALDANE—Mechanistic model of a random phenomenon. M. KALECKI—A study of large sample test criteria through properties of efficient estimates. C. RADHAKRISHNA RAO—A method of fractile graphical analysis. P. C. MAHALANOBIS—On some properties of error area in the fractile graph method. K. TAKEUCHI—Some limit distributions connected with fractile graphical analysis. J. SETHURAMAN—Some limit theorems in regression theory. K. R. PARTHASARATHY and P. K. BHATTACHARYA.

Series A, Vol. 23, Part 2

An elementary theory of the Brownian motion random function. J. KAMPÉ DE FÉRIET—On testing the mean of a discrete linear process. K. R. PARTHASARATHY—On a problem of Bartholomew in life testing. P. S. SWAMY and S. A. D. C. DOSS—Studentization of two-stage sample means from normal populations with unknown common variance. HAROLD RUBEN—Testing for the single outlier in a regression model. K. S. SRIKANTAN—Minimal sufficient statistics for the balanced incomplete block design under an Eisenhart model II. DAVID L. WEEKS and FRANKLIN A. GRAYBILL—Asymptotic bias and variance of ratio estimates in generalized power series distributions and certain applications. G. P. PATIL—Method of constructing two mutually orthogonal latin squares of order $3n+1$. P. KESAVA MENON—Combinatorial arrangements analogous to orthogonal arrays. C. RADHAKRISHNA RAO—On the structure and combinatorial properties of certain semi-regular group divisible designs. W. S. SARAF—A factorial approach to construction of true cost of living index and its application in studies of changes in national income. K. S. BANERJEE—Generation of random permutations of given number of elements using random sampling numbers. C. RADHAKRISHNA RAO.

The subscription price per volume for each of the series is: U.S.A. and Canada, $10.00; Other Countries, £3

The subscription price per volume for both the series taken together is:
U.S.A. and Canada, $18.00; Other Countries, £5. 6s.

Subscription is payable in advance to:

THE SECRETARY

Statistical Publishing Society, 204/1 Barrackpore Trunk Road, Calcutta–35, India

ROYAL STATISTICAL SOCIETY

THE JOURNAL OF THE ROYAL STATISTICAL SOCIETY is published in two series: SERIES A (GENERAL), four issues a year, annual subscription £4. 4*s*. post free, U.S. $12: price each part £1. 5*s*.; SERIES B (METHODOLOGICAL), two issues a year, annual subscription £3. 2*s*. post free, U.S. $9: price each part £2.

SERIES A (GENERAL), VOL. 124, PART 2, 1961

The Role of Models in Operational Research (with Discussion). By K. D. TOUCHER—Research on Mail Surveys (with Discussion). By CHRISTOPHER SCOTT—The International Distribution of Income. By S. ARDIC and A. T. PEACOCK—The Analysis of Survey Data on an Electronic Computer. By H. R. SIMPSON—A Practical Demonstration of a Needle Experiment designed to give a Number of Concurrent Estimates of π. By B. C. KAHAN—A Note on the Paper by W. E. Thomson on 'ERNIE—a Mathematical and Statistical Analysis'. By EVE BOFINGER and V. J. BOFINGER—The Recent Controversy over Resale Price Maintenance. By G. BORTS—Reviews—Current Notes—Obituary: E. C. Fieller—Book List.

SERIES B (METHODOLOGICAL), VOL. 23, PART 2, 1961

A Test of Homogeneity of Means under Restricted Alternatives (with Discussion). By D. J. BARTHOLOMEW—Optimal Programmes for Varietal Selection (with Discussion). By R. N. CURNOW—Some Stochastic Processes with Absorbing Barriers. By T. L. SAATY—Some Sequential Tests using Range. By W. G. GILCHRIST—First Emptiness in a Finite Dam. By B. WEESAKUL—The Construction of Optimal Designs for the One-Way Classification Analysis of Variance. By J. AITCHISON—Some Simple Wear-Dependent Renewal Processes. By A. MERCER—An Exact Distribution of the Behrens–Fisher–Welch Statistic for Testing the Difference between the Means of Two Normal Populations with Unknown Variances. By W. E. RAY and A. E. N. T. PITMAN—Close Queueing Systems: A Generalisation of the Machine Interference Model. By F. BENSON and G. GREGORY—Testing for a Jump in the Spectral Function. By E. J. HANNAN—Moments of a Truncated Bivariate Normal Distribution. By S. ROSENBAUM—On Quadratic Estimates of the Interclass Variance for Unbalanced Designs. By R. S. READ—Mixtures of Geometric Distributions. By H. E. DANIELS—Prediction by Exponentially Weighted Moving Averages and Related Methods. By D. R. COX—The Randomization Bases of the Problem of the Amalgamation of Weighted Means. By D. E. BARTON and C. L. MALLOWS—Estimating from Quantiles in Destructive Testing. By H. P. STOUT and F. STERN—A Note on Maximum-Likelihood in the Case of Dependent Random Variables. By S. D. SILVEY—Some Properties of the Asymptotic Variance of the Sample Quantiles and Mid-Ranges. By P. K. SEN—Similarities between Likelihoods and Associated Distributions *a Posteriori*. By D. A. SPROTT—A Method of Maximum Likelihood Estimation. By F. S. G. RICHARDS—The Intervals between Regular Events Displaced in Time by Independent Random Deviations of Large Dispersal. By T. LEWIS—An Approach to the Scheduling of Jobs on Machines. By E. PAGE.

ROYAL STATISTICAL SOCIETY, 21 BENTINCK STREET, LONDON, W.1

APPLIED STATISTICS

A JOURNAL OF THE ROYAL STATISTICAL SOCIETY

VOLUME X, NO. 1 CONTENTS MARCH 1961

Some Industrial Applications of Multivariate Analysis. B. A. M. THOMAS.

Measurements and Standards. FREDA CONWAY.

Some Notes on the Precision of the Gradient of an Estimated Straight Line. DONALD G. BEECH.

An Introduction to Dynamic Programming. M. G. SIMPSON.

Dynamic Programming and Decision Theory. D. V. LINDLEY.

A Method of Allowing for 'Not-at-home' Bias in Sample Surveys. D. J. BARTHOLOMEW.

Notes and Comments

Meetings of Sections and Local Groups of the Royal Statistical Society.

BOOK REVIEWS.

APPLIED STATISTICS is published three times per year. The annual subscription is £1. 10s. (U.S.A. and Canada $5.00). Single copies 13s. post free. Orders should be sent to

Oliver and Boyd Ltd., Tweeddale Court, 14 High Street, Edinburgh, 1

INTERNATIONAL JOURNAL OF ABSTRACTS STATISTICAL THEORY AND METHOD

A Journal of the International Statistical Institute

The aim of this journal of abstracts is to give complete coverage of published papers in the field of statistical theory (including associated aspects of probability and other mathematical methods) and new published contributions to statistical method.

All contributions in the following five journals—being wholly devoted to this field—are abstracted: *Annals of Mathematical Statistics; Biometrika; Journal Royal Statistical Society (Series B); Bulletin of Mathematical Statistics; Annals Institute of Statistical Mathematics;* and a further group of six journals are abstracted on a virtually complete basis as follows: *Biometrics; Metrika; Metron; Review International Statistical Institute; Technometrics; Sankhyā.* There are about 250 other journals partly devoted to statistical theory and method from which the appropriate papers are abstracted.

The abstracts are about 400 words long—the recommendation of UNESCO for the 'long' abstract service: They are in the English language although the original language of the paper is noted on the abstract together with the name of the abstractor. In addition the address of the author(s) are given in detail to facilitate contact in order to request an off-print or obtain further details. The journal is published quarterly and contains approximately 1000 abstracts per year.

A scheme of classification has been developed for the abstracts that is flexible and facilitates the transfer of code numbers to punched-cards. A unique aspect of this journal is that the pages are colour-tinted according to the main sections of classification. This method of colour-coding the pages provides a distinctive and powerful visual aid in the identification of abstracts in whatever manner the journal is filed for reference.

Annual Subscription	£5 (U.S.A. and Canada $16.00)
Single Number	£1. 10*s.* (U.S.A. and Canada $4.50)

OLIVER AND BOYD LTD.

Tweeddale Court, 14 High Street, Edinburgh, 1

TECHNOMETRICS

A Journal of Statistics for the Physical, Chemical and Engineering Sciences

CONTENTS

VOL. 3, NO. 2 — MAY 1961

General Considerations in the Analysis of Spectra. G. M. JENKINS.
Mathematical Considerations in the Estimation of Spectra. EMANUEL PARZEN.
Discussion, Emphasizing the Connection Between Analysis of Variance and Spectrum Analysis. JOHN W. TUKEY.
Some Comments on Spectral Analysis of Time Series. N. R. GOODMAN.
Comments on the Discussions of Messrs Tukey and Goodman. G. M. JENKINS and EMANUEL PARZEN.
Spectral Analysis Combining a Bartlett Window with an Associated Inner Window. THOMAS H. WONNACOTT.
Frequency Response from Stationary Noise: Two Case Histories. N. R. GOODMAN, S. KATZ, B. H. KRAMER and M. T. KUO.
The Modified Gauss–Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares. H. O. HARTLEY.
On the Possibility of Improving the Mean Useful Life of Items by Eliminating Those with Short Lives. G. S. WATSON and W. R. WELLS.
The Optimum Allocation of Spare Components in System. DONALD F. MORRISON.
Use of Tables of Percentage Points of Range and Studentized Range. H. LEON HARTER.
Book Review. C. DANIEL.
Statistical Programs for High Speed Computers.
Notices.

VOL. 3, NO. 3 — AUGUST 1961

The 2^{k-p} Fractional Factorial Designs. G. E. P. BOX and J. S. HUNTER.
Reduced Designs of Resolution Five. J. C. WHITWELL and G. K. MORBEY.
Partial Confounding in Fractional Replication. W. J. YOUDEN.
Finding New Fractions of Factorial Experimental Designs. R. E. FRY.
A Study of the Group Screening Method. G. S. WATSON.
Missing Values in Response Surface Designs. NORMAN R. DRAPER.
Use of Tables of Percentage Points of Range and Studentized Range. H. LEON HARTER.
The Optimum Allocation of Spare Components in Systems. DONALD F. MORRISON.
The Reliability of Components Exhibiting Cumulative Damage Effects. GEORGE H. WEISS.
An Analysis of Some Relay Failure Data from a Composite Exponential Population. R. R. PRAIRIE and B. OSTLE.
Applications of Truncated Distributions in Process Startups and Inventory Control. H. SMITH and D. W. GRACE.
Estimating the Poisson Parameter from Samples that are Truncated on the Right. A. CLIFFORD COHEN, Jr.
A General Simulation Programme for Batch Chemical Plants. J. DYSON, P. L. GOLDSMITH and J. S. M. ROBERTSON.
Book Reviews. M. J. R. HEALY and M. STONE.
Statistical Programs for High Speed Computers.
Notices.

Technometrics is published quarterly in February, May, August, and November. The annual non-member subscription rate is $8.00. To members of the American Statistical Association and the American Society for Quality Control the rate is $6.00. Checks should be made payable to *Technometrics* and addressed to *Technometrics*, Post Office Box 587, Benjamin Franklin Station, Washington 6, D.C.

The Annals of Mathematical Statistics

The Official Journal of the Institute of Mathematical Statistics

VOL. 32, NO. 3

CONTENTS

SEPTEMBER 1961

Oskar Anderson, 1887–1960. HERMAN WOLD—On Fiducial Inference. D. A. S. FRASER—A Central Limit Theorem for Partly Dependent Variables. H. J. GODWIN and S. K. ZAREMBA—Some Multivariate Chebyshev Inequalities with Extensions to Continuous Parameter Processes. Z. W. BIRNBAUM and ALBERT W. MARSHALL—Maximal Independent Stochastic Processes. C. B. BELL—On Markov Chain Potentials. JOHN G. KEMENY and J. LAURIE SNELL—Markov Chains with Absorbing States: A Genetic Example. G. A. WATTERSON—Estimation of the Spectrum. V. K. MURTHY—On a Coincidence Problem Concerning Particle Counters. LAJOS TAKÁCS—On the Ruin Problem of Collective Risk Theory. N. U. PRABHU—The Random Walk between a Reflecting and an Absorbing Barrier. B. WEESAKUL—On the Queueing Process M/G/1. C. R. HEATHCOTE—The Sequential Design of Experiments for Infinitely Many States of Nature. ARTHUR E. ALBERT—Analysis of a Class of PBIB Designs with more than Two Associate Classes. P. V. RAO—On a Locally Most Powerful Boundary Randomized Similar Test for the Independence of Two Poisson Variables. M. S. AHMED—Confidence Intervals from Censored Samples. MAX HALPERIN—Tests of Fit Based on the Number of Observations Falling in the Shortest Sample Spacings Determined by Earlier Observations. LIONEL WEISS—Some Nonparametric Median Procedures. V. P. BHAPKAR—On Certain Characteristics of the Distribution of the Latent Roots of a Symmetric Random Matrix under General Conditions. H. R. VAN DER VAART—The Distribution of Noncentral Means with Known Covariance. ALAN T. JAMES—Distribution of a Definite Quadratic Form for Non-Central Normal Variates. B. K. SHAH and C. G. KHATRI—Percentage Points and Modes of Order Statistics from the Normal Distribution. SHANTI S. GUPTA—NOTES: Expressing a Random Variable in Terms of Uniform Random Variables. G. MARSAGLIA—Generating Exponential Random Variables. G. MARSAGLIA—A Combinatorial Lemma for Complex Numbers. GLEN BAXTER—A Combinatorial Derivation of the Distribution of the Truncated Poisson Sufficient Statistic. T. CACOULLOS—A Note on Simple Binomial Sampling Plans. B. BRAINERD and T. V. NARAYANA—An Inequality for Balanced Incomplete Block Designs. V. N. MURTY—Third Order Rotatable Designs in Three Dimensions: Some Specific Designs. NORMAN R. DRAPER—Abstracts of Papers—News and Notices—Report of the Ithaca, New York Meeting—Publications Received.

VOL. 32, NO. 4

CONTENTS

DECEMBER 1961

An Approach to Time Series Analysis. EMANUEL PARZEN—Some Model I Problems of Selection. E. L. LEHMANN—Bayes Rules for a Common Multiple Comparisons Problem and Related Student-t Problems. DAVID B. DUNCAN—The Use of Least Favorable Distributions in Testing Composite Hypotheses. H. E. REINHARDT—Asymptotic Efficiency in Polynomial Estimation. PAUL G. HOEL—Maximum Likelihood Estimation of a Linear Functional Relationship. C. VILLEGAS—Sequential χ^2- and T^2-Tests. EDWARD J. JACKSON and RALPH A. BRADLEY—Estimating the Parameters of Negative Exponential Populations from One or Two Order Statistics. H. LEON HARTER—On the Two Sample Problem: A Heuristic Method for Constructing Tests. V. P. GODAMBE—A Nonparametric Test for the Problem of Several Samples. V. P. BHAPKAR—Distribution of the Anderson-Darling Statistic. PETER A. W. LEWIS—Errors in Discrimination. S. JOHN—On the Monotonic Character of the Power Functions of Two Multivariate Tests. S. N. ROY and W. F. MIKHAIL—The Moments of Elementary Symmetric Functions of the Roots of a Matrix in Multivariate Analysis. TITO A. MIJARES—Variance Components in the Unbalanced 2-Way Nested Classification. S. R. SEARLE—Some Main-Effect Plans and Orthogonal Arrays of Strength Two. SIDNEY ADDELMAN and OSCAR KEMPTHORNE—On a Geometrical Method of Construction of Partially Balanced Designs with Two Associate Classes. ESTER SEIDEN—On Some Methods of Construction of Partially Balanced Arrays. I. M. CHAKRAVARTI—Some Further Designs of Type $O:PP$. G. H. FREEMAN—Sufficiency in the Undominated Case. D. L. BURKHOLDER—On a Special Class of Recurrent Events. M. P. SCHÜTZENBERGER—Maximum Likelihood Characterization of Distributions. HENRY TEICHER—On the Distribution of First Significant Digits. ROGER S. PINKHAM—Markov Renewal Processes. Definitions and Preliminary Properties. RONALD PYKE—Markov Renewal Processes with Finitely Many States. RONALD PYKE—A Convexity Property in the Theory of Random Variables Defined on a Finite Markov Chain. H. D. MILLER—Limit Distribution in the Theory of Counters. G. SANKARANARAYANAN—The Transient Behavior of a Single Server Queuing Process with Recurrent Input and Gamma Service Time. LAJOS TAKÁCS—Efficient Estimation of a Regression Parameter for Certain Second Order Processes. CHARLOTTE T. STRIEBEL—Two Similar Queues in Parallel. J. F. C. KINGMAN—Queues with Batch Departures. I. F. G. FOSTER and K. M. NYUNT—NOTES: On the Chapman-Kolmogorov Equation. JACK KARUSH—A Generalization of a Theorem of Balakrishnan. N. DONALD YLVISAKER—The Opinion Pool. M. STONE—Correction Notes—Abstracts of Papers—News and Notices—Report of the President for 1961—Report of the Editor for 1961—Publications Received.

Subscription rate $12.00 *per year in the United States and Canada and* $10.00 *per year elsewhere*

ADDRESS ORDERS FOR SUBSCRIPTIONS AND BACK NUMBERS TO

PROFESSOR A. H. BOWKER, TREASURER, INSTITUTE OF MATHEMATICAL STATISTICS, DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY, STANFORD, CALIFORNIA, U.S.A.

BIOMETRIKA

BACK ISSUES

Volumes 44–48 (1957–1961)

These may be obtained from the BIOMETRIKA OFFICE, at the following prices, which include packing and postage:

£3. 10*s*. (or $10.00) per wrappered volume.
Bound volumes: £1. 8*s*. (or $4.00) extra per volume.
Binding cases: 14*s*. (or $2.00) each.

Cheques should be made payable to *Biometrika*, crossed 'a/c BIOMETRIKA TRUST' and sent to THE SECRETARY, BIOMETRIKA OFFICE, UNIVERSITY COLLEGE LONDON, GOWER STREET, LONDON, W.C. 1, to whom all orders for subscriptions and author's separates should also be sent (see inside front cover). The Secretary also has available for sale copies of many of the papers and tables published in earlier numbers of *Biometrika* (see pages (iv), (v) and (vi)).

Volumes 20A, 20B, 21–43 (1928–1956)

Messrs Wm. Dawson & Sons, Ltd., have been appointed agents for the sale of these volumes, which may be obtained from the address below at the following prices, *packing and postage extra:*

£5. 5*s*. (or $15.00) per wrappered volume.
Bound volumes: £1. 8*s*. (or $4.00) extra per volume.
Binding cases: 14*s*. (or $2.00) each.
Index (Subjects Vols. 1–37; Authors Vols. 1–43): 6*s*. (or $1.00).

Volumes 1–19 (1902–1927)

Messrs Wm. Dawson & Sons, Ltd., have been granted permission to reprint Vols. 1–19. These are ready for distribution, price £130 bound. Single vols. £7. 10*s*. each. Would librarians and others wishing to have copies please place their orders with:

WM. DAWSON & SONS, LTD.,
16 WEST STREET, FARNHAM, SURREY, ENGLAND

(*All rights reserved*)

BIOMETRIKA, Vol. 48, Parts 3 and 4

CONTENTS

PAGE

J. N. Darroch. The two-sample capture-recapture census when tagging and sampling are stratified . . 241
D. A. S. Fraser. The fiducial method and invariance 261
G. B. Wetherill. Bayesian sequential analysis 281
Byron Wm Brown, Jr. Some properties of the Spearman estimator in bioassay 293
Patrick Slater. Inconsistencies in a schedule of paired comparisons 303
George W. Morgenthaler. Some circular coverage problems 313
D. J. Bartholomew. Ordered tests in the analysis of variance 325
D. H. Young. Quota fulfilment using unrestricted random sampling 335
A. M. Walker. Large-sample estimation of parameters for moving-average models 343
Olav Reiersøl. Linear and non-linear multiple comparisons in logit analysis 359
J. G. Saw. Estimation of the normal population parameters given a type I censored sample . . . 367
C. P. Quesenberry and H. A. David. Some tests for outliers 379
J. F. C. Kingman. The ergodic behaviour of random walks 391
A. M. Kshirsagar. The goodness-of-fit of a single (non-isotropic) hypothetical principal component . . 397
S. John. On the evaluation of the probability integral of the multivariate t-distribution 409
J. P. Imhof. Computing the distribution of quadratic forms in normal variables 419
K. C. Chanda. Comparative efficiencies of methods of estimating parameters in linear autoregressive schemes 427
Frederick Mosteller and Cleo Youtz. Tables of the Freeman–Tukey transformations for the binomial and Poisson distributions 433

Miscellanea

C. F. Crouse. A non-null ranking model for a sequence of m alternatives 441
Gerald J. Glasser and Robert F. Winter. Critical values of the coefficient of rank correlation for testing the hypothesis of independence 444
J. G. Saw. The bias of the maximum likelihood estimates of the location and scale parameters given a type II censored normal sample 448
B. K. Kale. On the solution of the likelihood equation by iteration processes 452
A. Zinger. Detection of best and outlying normal populations with known variances 457
E. J. Burr and Gwenda Cane. Longest run of consecutive observations having a specified attribute . 461
D. Hogben, R. S. Pinkham and M. B. Wilk. The moments of the non-central t-distribution . . . 465
J. D. Biggers. Estimation of missing observations in split-plot experiments where whole-plots are missing or mixed-up 468
P. D. Finch. Deterministic customer impatience in the queueing system GI/M/1; a correction . . . 472

Corrigenda 474

Reviews

David Middleton. An Introduction to Statistical Communication Theory 477
Walter W. Garvin. Introduction to Linear Programming. 477
Howard Raiffa and Robert Schlaifer. Applied Statistical Decision Theory 478
Martin Shubik. Strategy and Market Structure. Competition, Oligopoly, and the Theory of Games . 479
R. L. Plackett. Regression Analysis 480
Samuel Goldberg. Probability. An Introduction. 481
Paul G. Hoel. Elementary Statistics 481
F. Yates. Sampling Methods for Censuses and Surveys 481
Henri Louis le Roy. Statistische Methoden der Populationsgenetik 482
J. W. L. Beament (Ed.) Symposia of the Society for Experimental Biology. No. XIV. Models and Analogues in Biology 482
P. G. Guest. Numerical Methods of Curve Fitting 482
E. Lukacs. Characteristic Functions 483
Albert Eagle. The Elliptic Functions as they should be 483
George E. Forsythe and Wolfgang R. Watson. Finite-difference Methods for Partial Differential Equations 484
George Boole. The Calculus of Finite Differences 484
W. H. Durfee. Fundamentals of College Algebra 484

Other Books Received 485

Printed in Great Britain at the University Press, Cambridge (Brooke Crutchley, University Printer)

AGE
241
261
281
293
303
313
325
333
343
359
367
379
391
397
409
419
427
433
441
444
448
452
457
461
465
468
472
474
477
477
478
479
480
481
481
481
482
482
482
483
483
484
484
484
485